

Instituto Superior de Economia e Gestão
Universidade Técnica de Lisboa

MODELOS ESTRUTURAIS

Abordagens Clássica e Bayesiana

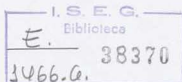
por

Maria de Fátima Ferrão

Dissertação apresentada como requisito
parcial para a obtenção de grau de
Mestre em Matemática Aplicada
à Economia e Gestão

JUNHO 1992

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO
UNIVERSIDADE TÉCNICA DE LISBOA



RESERVADO

HA30.3 F47
1992

MODELOS ESTRUTURAIS - Abordagens Clássica e Bayesiana

Dissertação apresentada como requisito parcial para a
obtenção do grau de mestre em Métodos Matemáticos
Aplicados à Economia e Gestão .

Orientador: Professor Doutor Daniel Muller

Maria de Fátima N. Urbano Lopes Ferrão

Junho de 1992



AGRADECIMENTOS

Ao Professor Doutor Daniel Muller, orientador desta dissertação, pelo constante estímulo e todas as sugestões que me foi dando ao longo da sua elaboração.

À Professora Doutora Antónia Turkman pela sua valiosa ajuda no que se refere à segunda parte deste trabalho.

Ao Professor Jeff Harrison do Departamento de Estatística da Universidade de Warwick pelos momentos que me dispensou aquando da sua visita ao I.S.E.G. (3ª Conferência do CEMAPRE), bem como, os vários artigos que, prontamente, me enviou após a sua chegada a Warwick.

A todos aqueles que, directa ou indirectamente, contribuíram para a realização deste trabalho, em especial à minha colega Margarida Chinita pela sua amizade e encorajamento sempre presentes.

1. INTRODUÇÃO

Principais Tópicos

MODELOS EXISTENTES DE Acreditação Clínica

1. MODELOS EXISTENTES

- 1.1 Introdução
- 1.2 Modelo Externacional (ISO)
- 1.3 Componentes de Tredwell (2)
- 1.4 Componente Académico, I
- 1.5 Componente Clínica (2)

2. MODELOS EM DESENVOLVIMENTO

- 2.1 Introdução
- 2.2 Modelo Externacional
- 2.3 Componente de Apoio Médico

3. MODELO KATMAN

- 3.1 Introdução
- 3.2 Descrição da Parte de Clínica
- 3.3 Descrição Interpretativa e/ou resultados da Parte de Clínica
- 3.4 Descrição da Parte de Clínica

1.6 Adicionalmente, TODOS OS ERROS OU OMISSÕES QUE ESTA DISSERTAÇÃO APRESENTA SÃO DA TOTAL RESPONSABILIDADE DA SUA AUTORA.

4. CONCLUSÃO

- 4.1 Introdução
- 4.2 Descrição da Parte de Clínica

5. APLICAÇÃO DO MODELO DE KATMAN A MODELOS EXISTENTES

- 5.1 Introdução
- 5.2 Descrição da Parte de Clínica
- 5.3 Descrição da Parte de Clínica
- 5.4 Descrição da Parte de Clínica

6. MODELOS EXISTENTES COM VARIÁVEIS EXPLORATIVAS

- 6.1 VARIÁVEIS DE INTRODUÇÃO

ÍNDICE

1. INTRODUÇÃO	1
Primeira Parte	
MODELOS ESTRUTURAIS – Abordagem Clássica	
2. MODELOS ESTRUTURAIS	4
2.1 Introdução.	
2.2 Modelo Estrutural Geral (MEG).	
2.3 Componente de Tendência (μ_t).	
2.4 Componente Sazonal (γ_t).	
2.5 Componente Cíclica (φ_t).	
3. MODELOS EM ESPAÇO DE ESTADOS	15
3.1 Introdução.	
3.2 Modelo Linear Dinâmico.	
3.3 Representação de Alguns Modelos.	
4. FILTRO DE KALMAN	25
4.1 Introdução.	
4.2 Definição do Filtro de Kalman.	
4.3 Dedução, Interpretação e Propriedades do Filtro de Kalman.	
4.4 Previsão em N Passos.	
4.5 Alisamento.	
4.6 Estabilidade do Filtro de Kalman.	
5. ESTIMAÇÃO	41
5.1 Introdução.	
5.2 Estimação de Máxima Verosimilhança . Decomposição no Erro de Previsão.	
6. APLICAÇÃO DO FILTRO DE KALMAN A MODELOS UNIVARIADOS	49
6.1 Introdução.	
6.2 Estimação de Máxima Verosimilhança no Domínio do Tempo.	
6.3 Estimação a Partir da Forma Reduzida. Correlograma.	
6.4 Estimação do M.E.B. no Domínio da Frequência.	
7. MODELOS ESTRUTURAIS COM VARIÁVEIS EXPLICATIVAS. E VARIÁVEIS DE INTERVENÇÃO	58

Segunda Parte

MODELOS ESTRUTURAIS – Abordagem Bayesiana

8. MODELOS ESTRUTURAIS – ABORDAGEM BAYESIANA	62
8.1 Introdução.	
8.2 Princípios Básicos da Inferência Bayesiana.	
8.3 Distribuições à Priori.	
8.3.1 Distribuições à Priori não Informativas.	
8.3.2 Famílias Conjugadas de Distribuições.	
8.4 Estimação Bayesiana.	
8.5 Regiões de Confiança e Testes de Hipóteses Bayesianos.	
9. MODELOS ESTRUTURAIS BAYESIANOS – FORMULAÇÃO MATEMÁTICA	82
9.1 Modelo Normal Linear Dinâmico Multivariado.	
9.2 Modelo Normal Linear Dinâmico Univariado.	
9.2.1 Inferência Sobre a Variância das Observações, h_t .	
9.2.2 Especificação da Matriz Q_t – Uso de Factores de Desconto.	
10. CONTROLO E INTERVENÇÃO.	103
10.1 Controlo - Via factor de Bayes.	
10.2 Intervenção.	
10.2.1 Tratamento de uma Observação, y_t , Errática.	
10.2.2 Perturbação Aleatória Adicional na Equação de Evolução.	
11. CONCLUSÃO.	110



1. INTRODUÇÃO

O grande impulso no estudo de séries temporais deu-se essencialmente a partir dos anos 70, com os trabalhos de Box & Jenkins (1970), óptica "clássica" e, quase simultaneamente, surge em 1971 a óptica "bayesiana" dos modelos de previsão proposta por Harrison & Stevens.

São fundamentalmente dois os motivos que conduzem à modelação de uma sucessão cronológica. O primeiro é a necessidade de obter um processo descritivo do fenómeno em termos das suas componentes de interesse e o segundo tem a ver com o aspecto da predição de observações futuras. Em geral, o modelo que é usado para descrever e explicar o fenómeno é também utilizado para previsão.

Os modelos tradicionais de Box e Jenkins (1970/76) são baseados na teoria dos processos estocásticos estacionários e são representados por um modelo misto autoregressivo e de médias móveis (ARMA). Uma classe mais vasta de modelos capazes de descrever comportamentos não estacionários pode ser obtida, admitindo que uma sucessão pode ser constituída por um processo ARMA após diferenciação. Tal classe conduz à definição dos modelos integrados autoregressivos e de médias móveis (ARIMA).

Muito poucas séries económicas ou sociais são estacionárias e é por vezes difícil torná-las estacionárias. Em particular a diferenciação da série pode causar problemas de interpretação da estrutura ARIMA identificada.

Problemas de identificação e interpretação, assim como a restrição de parâmetros constantes nos modelos ARIMA motivaram a procura de metodologias alternativas.

Uma classe de modelos paramétricos para previsão de sucessões cronológicas foi proposta recentemente sob a designação de "Modelos Estruturais" (Harvey (1981), (1984), (1989); Harvey e Todd (1983)). Seguindo os trabalhos pioneiros de Engle (1978) e Nerlove e al (1979). Estes modelos consistem essencialmente em impor uma estrutura particular às componentes não observáveis da tradicional decomposição de uma sucessão: **tendência**, **sazonalidade**, **ciclo** e **componente aleatória** (perturbação ou ruído).

O objectivo deste trabalho é o de introduzir, de uma maneira simples e tão clara quanto possível os desenvolvimentos teóricos fundamentais relativos aos modelos estruturais na óptica da sua aplicação à previsão econométrica.

O tema é novo e bastante vasto pelo que os problemas a estudar foram previamente delimitados. Optou-se por centrar a atenção essencialmente nos modelos lineares gaussianos e na aplicação do **filtro de Kalman**, a esses modelos, considerando-se as duas abordagens, "clássica" de Harvey e "bayesiana" de Harrison e Stevens.

O filtro de Kalman é um algoritmo recursivo de estimação, que se pode aplicar a qualquer modelo, desde que este se apresente sob a forma de espaço de estados. A formulação de um modelo em

espaço de estados consiste num conjunto de duas equações, uma designada **equação das observações** que descreve o processo gerador das observações em função de um conjunto de variáveis, chamadas **variáveis de estado**, outra chamada **equação de evolução ou transição** que descreve o modo como as variáveis de estado evoluem no tempo.

Pressupõe-se que o leitor está familiarizado com a teoria estatística bem como com os princípios básicos relativos aos modelos de previsão. Alguns conhecimentos sobre os modelos ARIMA de Box-Jenkins ajudarão a compreender a amplitude dos modelos estruturais, bem como o papel unificador do filtro de Kalman dos vários processos estatísticos de previsão.

Dividiu-se este trabalho em duas partes. Uma primeira parte - Abordagem Clássica, e uma segunda parte - Abordagem Bayesiana. A primeira parte contém 7 capítulos. Porém os capítulos 2 e 3 da primeira parte, que introduzem os modelos estruturais bem como a sua formulação em "espaço de estados", constituem ponto de partida para ambas as abordagens. A representação de modelos em espaço de estados, constitui o aspecto fundamental para o tratamento dos modelos estruturais para sucessões cronológicas, incorporando o vector de estado as várias componentes, não-observáveis da série, tais como tendência, sazonalidade, ciclos e componente irregular. No capítulo 4 apresentam-se os aspectos fundamentais relativos à **filtragem de Kalman**. Tratando nos capítulos 5 e 6 os problemas relativos à estimação dos parâmetros do modelo, sendo abordadas as duas metodologias - estimação no domínio do tempo e estimação no domínio da frequência. No último capítulo da primeira parte, capítulo 7, considera-se a possibilidade de tornar os modelos estruturais mais completos através da introdução de variáveis explicativas e de intervenção permitindo assim, uma melhor explicação do fenómeno a estudar bem como melhorar a qualidade das previsões obtidas.

A segunda parte consta de três capítulos. No capítulo 8 faz-se uma introdução genérica aos princípios básicos da **inferência Bayesiana**. No capítulo 9 mostra-se como esses princípios podem ser aplicados ao caso particular dos modelos estruturais, focando em particular os problemas da estimação sequencial da variância das observações e da variância das perturbações na equação de evolução via **factores de desconto**. No capítulo 10 abordam-se os aspectos do controlo dos modelos lineares dinâmicos, via **factor de Bayes**, e ainda os problemas da detecção de observações erráticas ("outliers") e de alteração de estrutura.

Nas conclusões tecem-se algumas comparações entre as duas abordagens e apontam-se alguns aspectos importantes, relacionados com o tema e que não foram considerados neste trabalho.

2. MODELOS ESTRUTURAIS

2.1 Introdução

Os processos mistos integrados autoregressivos e de médias móveis (ARIMA) introduzidos por Box e Jenkins (1976) constituem uma larga classe de modelos de previsão para séries cronológicas univariadas. Na teoria tradicional de Box e Jenkins, o cronograma da série, as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) desempenham um papel fundamental para a especificação do modelo adequado ao estudo da série. Contudo aqueles elementos são por vezes pouco informativos, especialmente quando se trata de pequenas amostras, além de que certas dificuldades de interpretação podem surgir devido à necessidade de diferenciação da série, o que por vezes pode conduzir à especificação de modelos inapropriados. O problema da interpretação real da estrutura ARIMA identificada para uma série, assim como a restrição de parâmetros constantes, motivaram a procura de metodologias alternativas que melhor descrevessem o dinamismo do sistema gerador da série. Um processo alternativo consiste em formular modelos directamente em termos das suas componentes tradicionais, não - observáveis, de tendência, sazonalidade, ciclo e componente irregular que têm uma interpretação directa e implicam uma estratégia mais formal de selecção e formulação dos modelos. Esta nova metodologia devida essencialmente a A. Harvey da London School of Economics, inclui a nova classe de modelos chamados **Modelos Estruturais**.

Nos trabalhos de Nerlove e al (1979) cada uma das componentes da série são modelizadas a partir de processos ARIMA.

O grande desenvolvimento nos anos 80 dos Modelos Estruturais dá-se sobretudo devido à "descoberta" pelos estatísticos do algoritmo de Kalman [Kalman (1960)] até aqui usado fundamentalmente em Engenharia de Controlo.

A aplicação do Filtro de Kalman a modelos estruturais faz uso da sua representação em **espaço de estados ou modelos lineares dinâmicos (MLD)**.

Dois princípios fundamentais contribuem para a grande flexibilidade dos Modelos Estruturais, os princípios da "decomposição" e da "sobreposição". É bem conhecido que uma combinação linear de modelos lineares ainda é um modelo linear. Este princípio da "sobreposição" estende-se pois no sentido de que **uma combinação linear de modelos lineares dinâmicos ainda é um modelo linear dinâmico**.

Neste capítulo apresentar-se-ão as diferentes formas de modelização de cada uma das componentes dos modelos estruturais, apresentando-se nos capítulos 3 e 4 os aspectos fundamentais referentes à representação de modelos sob a forma de espaço de estados e ao Filtro de Kalman.

2.2 Modelo Estrutural Geral (MEG).

O modelo estrutural geral (MEG) é directamente formulado para uma série cronológica univariada (y_t), como uma soma de quatro componentes: **tendência** (μ_t), **sazonalidade** (γ_t), **cíclica** (φ_t) e **componente aleatória** (ϵ_t), também designada por **ruído** ou **perturbação das observações**. Sendo estas componentes não - observáveis.

$$(MEG) \quad y_t = \mu_t + \gamma_t + \varphi_t + \epsilon_t \quad ; \quad t=1,2,\dots,T \quad (2.1)$$

Esta equação é designada por **equação das medidas** ou **equação das observações**.

Para a componente aleatória admite-se que se comporta como um "ruído branco" Gaussiano ou seja :

$$\epsilon_t \sim N(0; \sigma_\epsilon^2), \quad t=1,2,\dots \quad ; \quad E[\epsilon_t \cdot \epsilon_s] = \begin{cases} 0 & \text{se } t \neq s \\ \sigma_\epsilon^2 & \text{se } t = s \end{cases} \quad t, s = 1,2,\dots$$

Se na série cronológica a componente cíclica (φ_t) não se encontra presente obtem-se o **modelo estrutural básico (MEB)** assim designado por Harvey & Todd (1983).

$$(MEB) \quad y_t = \mu_t + \gamma_t + \epsilon_t \quad ; \quad t=1,2,\dots,T \quad (2.2)$$

2.3 Componente de Tendência (μ_t)

Para o processo estocástico (μ_t), representando o nível da série temporal y_t sob estudo, podemos admitir uma evolução localmente linear,

$$\mu_t = \mu_{t-1} + \beta_{t-1} + u_t \quad (2.3)$$

$$\beta_t = \beta_{t-1} + \delta_t \quad (2.4)$$

onde :

β_t é o declive da tendência no instante t .

u_t e δ_t são perturbações aleatórias gaussianas independentes uma da outra e ao longo do tempo, de média nula e variâncias σ_u^2 e σ_δ^2 , i.e.,

$$u_t \sim N(0; \sigma_u^2) \quad ; \quad E[u_t \cdot u_s] = \begin{cases} 0 & \text{se } t \neq s \\ \sigma_u^2 & \text{se } t = s \end{cases}$$

$$\varepsilon_t \sim N(0; \sigma_\varepsilon^2) \quad ; \quad E[\varepsilon_t \cdot \varepsilon_s] = \begin{cases} 0 & \text{se } t \neq s \\ \sigma_\varepsilon^2 & \text{se } t = s \end{cases}$$

As equações (2.3) e (2.4) constituem a formulação para a tendência, conhecida por **Modelo de Crescimento Linear** devida a Harrison & Stevens (1976).

Se assumirmos o declive nulo o modelo acima para a tendência passa a ser descrito pela única equação dinâmica :

$$\mu_t = \mu_{t-1} + u_t \quad (2.5)$$

A equação (2.5) traduz a formulação mais simples para a tendência (localmente constante), conhecida por **passeio aleatório** ou **modelo estacionário**, também de acordo com Harrison & Stevens (1976).

Admitindo apenas a presença da componente de tendência na série, a equação das observações (2.1) reduz-se a

$$y_t = \mu_t + \varepsilon_t \quad (2.6)$$

onde ε_t tem as propriedades atrás descritas.

Considerando o modelo estacionário para a tendência (2.5) e a equação das observações (2.6) podemos escrever

$$\begin{aligned} \Delta \mu_t &= \mu_t - \mu_{t-1} = u_t \\ \Delta y_t &= u_t + \Delta \varepsilon_t \quad \text{ou} \quad (1-L)y_t = e_t + \theta e_{t-1} \end{aligned} \quad (2.7)$$

sendo Δ o operador de primeiras diferenças, relacionado com o operador de atraso " L " por $\Delta = 1-L$ e e_t o erro de previsão.

A expressão (2.7) pode obter-se usando o método de alisamento exponencial para estimar a tendência. Assim, designando por $\hat{\mu}_t$ essa estimativa, teremos :

$$\hat{\mu}_t = a y_t + (1-a)\hat{\mu}_{t-1} \quad \text{ou seja} \quad \dot{\mu}_t = \hat{\mu}_{t-1} + a(y_t - \hat{\mu}_{t-1}) \quad (2.8)$$

onde a é a constante de alisamento.

Para este modelo a previsão a um passo é dada por $\hat{y}_t(1) = \hat{\mu}_t$ para todo t , logo $\hat{y}_{t-1}(1) = \hat{\mu}_{t-1}$, donde o erro de previsão $e_t = y_t - \hat{y}_{t-1} = y_t - \hat{\mu}_{t-1}$ ou seja, tem-se $\hat{\mu}_{t-1} = y_t - e_t$, pelo que (2.8) pode escrever-se

$$\hat{\mu}_t = y_t - e_t + a e_t = y_t - (1-a) e_t \quad \text{ou} \quad y_t = \hat{\mu}_t + (1-a) e_t$$

e portanto

$$y_t - y_{t-1} = \hat{\mu}_t - \hat{\mu}_{t-1} + (1-a) e_t - (1-a) e_{t-1} = a e_t + e_t - a e_t - (1-a) e_t$$

ou

$$\Delta y_t = (1-L) y_t = e_t - \theta e_{t-1} \quad \text{com} \quad \theta = 1 - a$$

Assim, se os $\{e_t\}$ forem "ruído branco", o modelo estacionário, apesar de bastante simples, é equivalente a um ARIMA (0,1,1).

Para o modelo de crescimento linear usando o operador Δ podemos escrever:

$$\Delta \mu_t = \beta_{t-1} + u_t$$

$$\Delta \beta_t = \delta_t$$

consequentemente :

$$\Delta^2 \mu_t = \Delta \beta_{t-1} + \Delta u_t = \delta_{t-1} + \Delta u_t$$

Aplicando segundas diferenças a (2.6) e substituindo estes resultados vem:

$$\Delta^2 y_t = \Delta^2 \mu_t + \Delta^2 e_t = \delta_{t-1} + \Delta u_t + \Delta^2 e_t$$

ou equivalentemente :

$$\Delta^2 y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \tag{2.9}$$

o que nos permite concluir que y_t é um ARIMA (0,2,2), desde que os erros se comportem como um "ruído branco".

A expressão (2.9) pode obter-se usando o método de Holt para estimar o nível ou tendência, $\hat{\mu}_t$, e o seu declive, $\hat{\beta}_t$. O método de Holt usa o princípio da ponderação, e faz uso de duas constantes de alisamento, uma para a tendência e outra para o seu declive. As equações que definem as estimativas da tendência, declive e previsão são, respectivamente:

$$\hat{\mu}_t = a y_t + (1-a) (\hat{\mu}_{t-1} + \hat{\beta}_{t-1}), \quad 0 < a < 1 \quad (2.10)$$

$$\hat{\beta}_t = b (\hat{\mu}_t - \hat{\mu}_{t-1}) + (1-b) \hat{\beta}_{t-1}, \quad 0 < b < 1 \quad (2.11)$$

$$\hat{y}_t(1) = \hat{\mu}_t + \hat{\beta}_t \quad (2.12)$$

Em função do erro de previsão $e_t = y_t - \hat{y}_{t-1}(1) = y_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}$ podemos reescrever (2.10) e (2.11) como se segue:

$$\hat{\mu}_t = \hat{\mu}_{t-1} + \hat{\beta}_{t-1} + a e_t \quad (2.13)$$

$$\hat{\beta}_t = \hat{\beta}_{t-1} + a b e_t \quad (2.14)$$

Da expressão de e_t retira-se

$$y_t = \hat{\mu}_{t-1} + \hat{\beta}_{t-1} + e_t \quad (2.15)$$

$$y_{t-1} = \hat{\mu}_{t-2} + \hat{\beta}_{t-2} + e_{t-1} \quad (2.16)$$

Subtraindo (2.16) de (2.15) e usando (2.13) e (2.14) vem:

$$\begin{aligned} y_t - y_{t-1} &= \hat{\mu}_{t-1} - \hat{\mu}_{t-2} + \hat{\beta}_{t-1} - \hat{\beta}_{t-2} + e_t - e_{t-1} \\ &= \hat{\beta}_{t-1} + a e_{t-1} + e_t - e_{t-1} \\ &= \hat{\beta}_{t-2} + a b e_{t-1} + a e_{t-1} + e_t - e_{t-1} \\ &= y_{t-1} - \hat{\mu}_{t-2} - e_{t-1} + a b e_{t-1} + a e_{t-1} + e_{t-1} - e_{t-1} \end{aligned}$$

ou

$$y_t - 2 y_{t-1} = -\hat{\mu}_{t-2} - (2 - a b - a) e_{t-1} + e_t \quad (2.17)$$

Escrevendo (2.16) para $t = t - 2$ vem :

$$y_{t-2} = \hat{\mu}_{t-3} + \hat{\beta}_{t-3} + e_{t-2} \quad (2.18)$$

Somando (2.18) a (2.17) tem-se:

$$y_t - 2y_{t-1} + y_{t-2} = \hat{\mu}_{t-3} + \hat{\beta}_{t-3} + e_{t-2} - \hat{\mu}_{t-2} - (2 - a - b - a) e_{t-1} + e_t \quad (2.19)$$

como $\hat{\mu}_{t-2} - \hat{\mu}_{t-3} - \hat{\beta}_{t-3} = a e_{t-2}$, substituindo em (2.19) vem

$$y_t - 2y_{t-1} + y_{t-2} = e_t - (2 - a - a - b) e_{t-1} - (a - 1) e_{t-2} \quad \text{ou}$$

$$(1 - L)^2 y_t = \Delta^2 y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \quad \text{com } \theta_1 = 2 - a - a - b \text{ e } \theta_2 = a - 1$$

como se pretendia provar.

2.4 Componente Sazonal (γ_t).

A sazonalidade pode ser introduzida de diversas maneiras, sendo uma alternativa possível a apresentada por Harvey (1984a, appendix B) onde admite que os efeitos sazonais podem variar ao longo do tempo, considerando como processo gerador da componente sazonal, γ_t , o seguinte:

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad ; \quad t=1, 2, \dots, T \quad (2.20)$$

onde:

s é o número de períodos no ano ($s=12$ séries mensais, $s=4$ trimestrais, etc)

γ_{t-j} é o factor sazonal correspondente ao instante "t-j" ($j=0, 1, \dots, s-1$)

ω_t é o termo perturbação ou ruído associado à sazonalidade no instante t.

Admite-se que:

$$\omega_t \sim N(0; \sigma_\omega^2), \quad t=1, 2, \dots \quad ; \quad E[\omega_t \cdot \omega_s] = \begin{cases} 0 & \text{se } t \neq s \\ \sigma_\omega^2 & \text{se } t = s \end{cases}$$

Esta modelização não implica que a soma dos factores sazonais, γ_t , para s períodos consecutivos, seja zero. No entanto esta situação está garantida em termos do valor esperado da soma desses mesmos factores.

Escrevendo a expressão (2.20) em termos do operador de atraso, "L", vem :

$$(1 + L + \dots + L^{s-1}) \gamma_t = S(L) \gamma_t = \omega_t \quad ; \quad t=1,2,\dots,T \quad (2.21)$$

visto que :

$$\Delta_s = 1 - L^s = (1 + L + \dots + L^{s-1}) (1-L) = S(L) \Delta$$

podemos ainda escrever:

$$\Delta_s \gamma_t = (1-L) \omega_t \quad \text{ou} \quad S(L) \gamma_t = \omega_t$$

Um modo alternativo de modelizar a componente sazonal, segundo Harrison e Akram (1983) é a partir de uma série trigonométrica de senos e cossenos, isto é, a partir das suas harmónicas mais significativas .

Assim, o efeito sazonal no instante t , γ_t , é obtido pela soma das contribuições individuais, γ_{jt} , de cada uma das j harmónicas na frequência λ_j correspondente :

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{jt} \quad ; \quad t = 1, 2, \dots, T \quad (2.22)$$

$$\gamma_{jt} = \alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t + \omega_{jt} \quad (2.23)$$

Sendo $\lambda_j = \frac{2\pi j}{s}$, $j = 1, \dots, s/2$ (admitindo-se s par) e α_j e β_j os coeficientes associados à harmónica "j". Então γ_{jt} e a sua componente associada γ_{jt}^* são obtidas a partir do sistema dinâmico :

$$\begin{bmatrix} \gamma_{jt} \\ \gamma_{jt}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega_{jt}^* \end{bmatrix} \quad ; \quad t=1, \dots, \left(\frac{s}{2}-1\right) \quad (2.24)$$

$$\text{e } \gamma_{jt} = (\cos \lambda_j) \gamma_{j,t-1}^* + \omega_{jt} \quad ; \quad j = s/2 \quad (2.25)$$

Os ruídos ω_{jt} e ω_{jt}^* , $j = 1, 2, \dots, s/2$, admitem-se normais, independentes de média nula e de variância comum σ_ω^2 por forma a que o factor sazonal só esteja dependente de um único parâmetro, tal como a representação através de factores (2.20).

Para obter (2.24) vamos partir de (2.23) e assumir, sem perda de generalidade, que o ruído ω_{jt} é

identicamente nulo. Assim, podemos escrever :

$$\begin{aligned}\gamma_{jt} &= \alpha_j \cos [\lambda_j (t-1) + \lambda_j] + \beta_j \sin [\lambda_j (t-1) + \lambda_j] \\ &= \alpha_j [\cos \lambda_j (t-1) \cos \lambda_j - \sin \lambda_j (t-1) \sin \lambda_j] + \beta_j [\sin \lambda_j (t-1) \cos \lambda_j + \cos \lambda_j (t-1) \sin \lambda_j] \\ &= [\alpha_j \cos \lambda_j (t-1) + \beta_j \sin \lambda_j (t-1)] \cos \lambda_j + [-\sin \lambda_j (t-1) + \beta_j \cos \lambda_j (t-1)] \sin \lambda_j\end{aligned}$$

Definindo $\gamma_{jt}^* = -\alpha_j \sin \lambda_j t + \beta_j \cos \lambda_j t$ vem

$$\gamma_{jt} = \gamma_{j(t-1)} \cos \lambda_j + \gamma_{j(t-1)}^* \sin \lambda_j \quad (2.26)$$

Procedendo de modo análogo para γ_{jt}^* obter-se-ia

$$\gamma_{jt}^* = -\gamma_{j(t-1)} \sin \lambda_j + \gamma_{j(t-1)}^* \cos \lambda_j \quad (2.27)$$

O processo inicializar-se-á com :

$$\begin{aligned}\gamma_{j0} &= \alpha_j \cos 0 + \beta_j \sin 0 = \alpha_j \\ \gamma_{j0}^* &= -\alpha_j \sin 0 + \beta_j \cos 0 = \beta_j\end{aligned}$$

No caso particular do **Modelo Estrutural Básico (MEB)** descrito por (2.2), assumindo para a tendência o modelo de crescimento linear descrito por (2.3) e (2.4) e para a sazonalidade o processo gerado por (2.20), atendendo a que

$$\Delta \mu_t = \beta_{t-1} + u_t ; \quad \Delta \beta_{t-1} = \delta_{t-1} \text{ e } S(L) \gamma_t = \omega_t$$

podemos escrever :

$$y_t = \frac{u_t}{\Delta} + \frac{\delta_{t-1}}{\Delta^2} + \frac{\omega_t}{S(L)} + \epsilon_t ; \quad t=1,2,\dots,T \quad (2.28)$$

onde os dois primeiros termos do 2º membro de (2.28) representam a tendência e o terceiro a sazonalidade. Assim, o modelo cai dentro da classe de modelos **UCARIMA (Unobserved - Component - Autorregressive - Integrated - Moving Average)** [Nerlove e al (1979), Box e al (1978) e Engle (1978)].

Multiplicando (2.28) pelo operador das 1^{as} diferenças, Δ , e pelo operador de diferenças sazonais, Δ_s , afim de o tornar estacionário obtém-se :

$$\Delta \Delta_s y_t = \Delta_s u_t + S(L) \delta_{t-1} + \Delta^2 \omega_t + \Delta \Delta_s \epsilon_t ; t = s+2, \dots, T \quad (2.29)$$

O 2.^o membro de (2.29) é um $MA(s+1)$ com restrições não lineares nos parâmetros os quais são funções das variâncias dos ruídos.

Assim, (2.29) representa a **forma reduzida ou canónica** do modelo estrutural básico, veja-se Engle (1978) e Nerlove e al (1979).

2.5 Componente Cíclica (φ_t).

Na modelização da maior parte das séries económicas é frequente a necessidade de introdução de uma componente cíclica representando um processo estocástico estacionário com movimento oscilatório em torno da sua tendência.

A especificação da componente cíclica pode ser feita com um processo autoregressivo como em Kitagawa (1981) ou como um ciclo estocástico, Harvey (1984_b) e (1989).

Segundo Harvey, seja φ_t a componente cíclica que representa uma função cíclica com frequência λ_c , medida em radianos. O período do ciclo é $2\pi/\lambda_c$. A componente cíclica pode ser expressa como uma onda sinoidal ou cosinoidal com parâmetros adicionais representando a amplitude e a fase. Assim, se for usada uma onda cosinoidal teremos:

$$\varphi_t = A \cos(\lambda_c t - \theta) ; t = 1, 2, \dots, T \quad (2.30)$$

onde A é a amplitude e θ a fase.

Uma formulação mais conveniente pode ser obtida se se considerar a componente cíclica como uma mistura de senos e cosenos. A amplitude e a fase são substituídas por dois novos parâmetros α e β , e

$$\varphi_t = \alpha \cos \lambda_c t + \beta \sin \lambda_c t \quad (2.31)$$

onde $(\alpha^2 + \beta^2)^{1/2}$ é a amplitude e $\text{tg}^{-1}(\beta/\alpha)$ é a fase.

A componente cíclica definida em (2.31) pode ser construída recursivamente de uma forma idêntica à da sazonalidade:

$$\begin{bmatrix} \varphi_t \\ \varphi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \varphi_{t-1} \\ \varphi_{t-1}^* \end{bmatrix} ; \quad t=1,2,\dots,T \quad (2.32)$$

onde $\varphi_0 = \alpha$ e $\varphi_0^* = \beta$.

A componente cíclica pode tornar-se estocástica pela adição em (2.32) de dois ruídos k_t e k_t^* :

$$\begin{bmatrix} \varphi_t \\ \varphi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \varphi_{t-1} \\ \varphi_{t-1}^* \end{bmatrix} + \begin{bmatrix} k_t \\ k_t^* \end{bmatrix} ; \quad t=1,2,\dots,T \quad (2.33)$$

Para que o modelo seja identificável é necessário assumir que os dois ruídos têm a mesma variância ($\text{var}[k_t] = \text{var}[k_t^*] = \sigma_k^2$) ou que são não correlacionados.

Por razões de parcimónia, na prática, geralmente as duas condições são impostas.

No modelo estocástico da componente cíclica descrito em (2.33) pode introduzir-se maior flexibilidade, introduzindo um **factor de amortecimento**, ρ_c .

Assim, a especificação mais geral da componente cíclica estocástica é:

$$\begin{bmatrix} \varphi_t \\ \varphi_t^* \end{bmatrix} = \rho_c \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \varphi_{t-1} \\ \varphi_{t-1}^* \end{bmatrix} + \begin{bmatrix} k_t \\ k_t^* \end{bmatrix} ; \quad t=1,2,\dots,T \quad (2.34)$$

com k_t e k_t^* ruídos gaussianos continuamente independentes de média nula e variância iguais a σ_k^2 , e $0 \leq \rho_c \leq 1$. Se ρ_c é estritamente menor que 1 o processo gerador de φ_t é estacionário e é equivalente a um processo misto autoregressivo e de médias móveis de ordem (2,1), com fortes restrições nos parâmetros.

Com efeito desenvolvendo (2.34) podemos escrever :

$$\varphi_t = \rho_c \cos \lambda_c \varphi_{t-1} + \rho_c \sin \lambda_c \varphi_{t-1}^* + k_t$$

$$\varphi_t^* = -\rho_c \sin \lambda_c \varphi_{t-1} + \rho_c \cos \lambda_c \varphi_{t-1}^* + k_t^*$$

Substituindo φ_{t-1}^* , na expressão de φ_t obtém-se :

$$\begin{aligned}\varphi_t &= \rho_c \cos \lambda_c \varphi_{t-1} + \rho_c \sin \lambda_c [-\rho_c \sin \lambda_c \varphi_{t-2} + \rho_c \cos \lambda_c \varphi_{t-2}^* + k_{t-1}^*] + k_t \\ &= \rho_c \cos \lambda_c \varphi_{t-1} - \rho_c^2 \sin^2 \lambda_c \varphi_{t-2} + \rho_c^2 \sin \lambda_c \cos \lambda_c \varphi_{t-2}^* + \rho_c \sin \lambda_c k_{t-1}^* + k_t\end{aligned}$$

Somando e subtraindo $\rho_c \cos \lambda_c \varphi_{t-1}$ ao 2º membro da expressão anterior e substituindo na parcela negativa φ_{t-1} pela sua expressão anula-se o termo em φ_{t-2}^* e obtém-se :

$$\varphi_t = 2 \rho_c \cos \lambda_c \varphi_{t-1} - \rho_c^2 \varphi_{t-2} + k_t - \rho_c \cos \lambda_c k_{t-1} + \rho_c \sin \lambda_c k_{t-1}^* \quad (2.35)$$

ou usando o operador de atraso L , teremos

$$[1 - 2 \rho_c \cos \lambda_c \cdot L + \rho_c^2 \cdot L^2] \varphi_t = (1 - \rho_c \cos \lambda_c \cdot L) k_t + (\rho_c \sin \lambda_c \cdot L) k_t^* \quad (2.36)$$

A expressão (2.36) apresenta pois a forma de um ARMA (2,1).

Os zeros do polinómio autoregressivo, $(1 - 2 \rho_c \cos \lambda_c \cdot L + \rho_c^2 \cdot L^2 = 0)$ são complexos conjugados da forma $r_{1,2} = \frac{1}{\rho_c} (\cos \lambda_c \pm i \sin \lambda_c)$, sendo $|r_{1,2}| = \frac{1}{\rho_c}$ e como $0 \leq \rho_c \leq 1$, implica $|\frac{1}{\rho_c}| \geq 1$. Se ρ_c for estritamente menor que 1 a estacionaridade do processo é assegurada.

Harvey (1985) no seu trabalho aplica os modelos estruturais a várias séries macroeconómicas dos Estados Unidos (PNB, taxa de desemprego, preços no consumidor, etc), tendo estimado o modelo com e sem a restrição de $\sigma_k^2 = \sigma_{k^*}^2$.

Nalgumas aplicações pode ser necessário introduzir a componente cíclica na modelização da tendência, ou seja formaliza-se um modelo com tendência cíclica da forma :

$$y_t = \mu_t + \epsilon_t \quad (2.37)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \varphi_{t-1} + \eta_t \quad (2.38)$$

$$\beta_t = \beta_{t-1} + \delta_t \quad (2.39)$$

com φ_t descrito por (2.34).

3. MODELOS EM ESPAÇO DE ESTADOS.

3.1 Introdução

A formulação de modelos em **espaço de estados** surgiu com o desenvolvimento da teoria dos sistemas de controlo e é baseada na propriedade dos Sistemas Markovianos que estabelece a independência do futuro do processo em relação ao seu passado, dado o estado presente. Nestes sistemas o **estado** do processo condensa toda a informação do passado necessária para a predição do futuro.

A representação de modelos em espaço de estados reveste-se de particular importância devido à abertura para a utilização da **filtragem de Kalman** como um método de estimação, predição e alisamento a partir do conhecimento das observações y_t .

3.2 MODELO LINEAR DINÂMICO

A formulação dinâmica das componentes não observáveis descrita no capítulo anterior sendo linear é compatível com a representação sob a forma de espaço de estados via um **Modelo Linear Dinâmico (MLD)**.

Um **Modelo Linear Dinâmico** consiste num conjunto de duas equações dinâmicas estocásticas, uma descrevendo a evolução das observações (**equação de medida ou das observações**) e outra descrevendo a evolução dos parâmetros do sistema (**equação de transição ou equação de estado**).

A formulação geral em espaço de estados aplica-se a séries multivariadas y_t , contendo N elementos. Os valores observados da série cronológica admitem-se dependentes de m variáveis de estado, não observáveis que constituem o **vector de estado** θ_t ($m \times 1$). A relação entre y_t e θ_t é traduzida pela **equação de medida ou das observações**

$$y_t = F_t \theta_t + \epsilon_t \quad ; \quad t=1,2,\dots,T \quad (3.1)$$

onde:

F_t é uma matriz ($N \times m$), conhecida.

θ_t é o vector de estado ($m \times 1$), desconhecido.

ϵ_t é o vector ($N \times 1$) dos erros das observações que se admite se comporta como um "ruído branco",

$$E[\epsilon_t] = 0 \text{ para qualquer } t; \quad E[\epsilon_t \cdot \epsilon_k^T] = H_t \quad \text{e} \quad E[\epsilon_t \cdot \epsilon_k^T] = 0 \text{ se } t \neq k$$

No caso de séries cronológicas univariadas, $N=1$, a equação das observações escreve-se:

$$y_t = f_t^T \theta_t + \epsilon_t \quad ; \text{ com} \quad (3.2)$$

$$E[\epsilon_t] = 0, \quad \text{var}[\epsilon_t] = h_t \quad ; t=1,2,\dots,T.$$

Os elementos de θ_t , como se disse, em geral são não observáveis, contudo admite-se conhecer o modo como o vector de estado evolui no tempo, e assume-se que os estados são gerados por um processo de Markov de 1.ª ordem:

$$\theta_t = G_{t-1} \theta_{t-1} + R_t \eta_t \quad ; t = 1,2,\dots,T \quad (3.3)$$

onde G_t (mxm) e R_t (mxq) representam, respectivamente, a matriz de transição e a matriz relativa à parte aleatória, η_t é um vector (qx1) das perturbações(ruídos) associadas ao estado, que mais uma vez se assume se comporta como um "ruído branco"

$$E[\eta_t] = 0, \quad Q_t = E[\eta_t \eta_t^T] \text{ e } E[\eta_t \eta_k^T] = 0 \quad \text{para } t \neq k.$$

A equação (3.3) é a equação de transição. A inclusão da matriz R_t associada à parte aleatória é de certo modo arbitrária.

O MLD descrito pelas equações (3.1) e (3.3) contém duas componentes aleatórias distintas, i.e, ϵ_t (associada às observações) e η_t (associada ao estado). A especificação do MLD fica completa com as duas hipóteses seguintes:

- (1) - O vector de estado inicial, θ_0 , tem média a_0 e matriz de covariâncias P_0 , i.e.,

$$E[\theta_0] = a_0 \text{ e } E[(\theta_0 - a_0)(\theta_0 - a_0)^T] = P_0 \quad (3.4)$$

- (2) - As perturbações ϵ_t e η_t são não correlacionadas em todos os períodos de tempo e não correlacionadas com o estado inicial, i.e.,

$$E[\epsilon_s \cdot \eta_t^T] = 0 \quad ; \quad s, t = 1, 2, \dots, T \quad (3.5)$$

$$E[\epsilon_t \cdot \theta_0^T] = 0; \quad E[\eta_t \cdot \theta_0^T] = 0 \quad t=1,2,\dots,T \quad (3.6)$$

As matrizes F_t e H_t na equação das observações e as matrizes G_t , R_t e Q_t na equação de transição serão referidas como **matrizes do sistema**. Salvo referência em contrário admitir-se-á que são não estocásticas. Assim embora possam variar no tempo, fá-lo-ão de uma maneira determinística. Como já se disse, o sistema é linear e, para qualquer valor de t , y_t pode ser expresso como uma combinação linear das perturbações presentes e passadas, ϵ_t e η_t e o vector de estado inicial θ_0 .

Se as matrizes do sistema, F_t , H_t , G_t , R_t e Q_t , não mudam ao longo do tempo, o modelo diz-se, **invariante no tempo** ou **homogéneo no tempo**.

O **MLD** para qualquer processo, y_t , fica totalmente definido uma vez conhecidas as cinco matrizes do sistema, por isso é comum definir o **MLD** através do conjunto

$$M_t = \{ F_t, H_t, G_t, R_t, Q_t \}$$

conhecido como **vector de caracterização do MLD** [Harrison & Stevens (1976)].

Qualquer modelo linear pode facilmente formular-se sob a forma de espaço de estados. Vejamos alguns exemplos, com particular destaque para os modelos estruturais que tratámos no capítulo anterior.

3.3 - REPRESENTAÇÃO DE ALGUNS MODELOS

A representação de um modelo sob a forma de espaço de estados não é única, sendo sempre possível encontrar outras equivalentes. No entanto o objectivo da formulação em espaço de estados é que o vector de estado, θ_t , contenha toda a informação do sistema no instante t , usando o menor número possível de elementos. Uma representação em espaço de estados que minimiza a dimensão do vector de estado é dita - **realização mínima**. Uma preocupação, sempre presente, deve ser pois a de obter realizações mínimas. Contudo, isso não implica necessariamente a unicidade da representação. De facto a representação única é uma excepção e não uma regra. Como facilmente se pode ver definindo uma matriz arbitrária não-singular, B ($m \times m$), e considerando um novo vector de estado $\theta_t^* = B \theta_t$, premultiplicando a equação de transição original (3.3) por B vem

$$\theta_t^* = G_{t-1}^* \theta_{t-1}^* + R_t^* Q_t \quad (3.7)$$

onde

$$G_t^* = B G_{t-1} B^{-1} \quad \text{e} \quad R_t^* = B R_t \quad (3.8)$$

A correspondente equação de medida é:

$$y_t = F_t^* \theta_t^* + \epsilon_t \quad \text{com} \quad F_t^* = F_t B^{-1} \quad (3.9)$$

(1) – Modelo de Regressão Linear

Suponhamos que a variável observada y_t depende linearmente de k variáveis explicativas, X_{1t} , X_{2t} , ..., X_{kt}

$$y_t = \alpha_{1t} X_{1t} + \alpha_{2t} X_{2t} + \dots + \alpha_{kt} X_{kt} + \epsilon_t \quad (3.10)$$

onde os coeficientes α_{it} ($i=1,2,\dots,k$) podem variar, por exemplo, de acordo com um passeio aleatório com perturbações η_{1t} , η_{2t} , ..., η_{kt} de variância σ_η^2 . As perturbações, ϵ_t , associadas às observações, supõem-se ser, como habitual, um "ruído branco".

Se considerarmos $F_t = [X_{1t} \ X_{2t} \ \dots \ X_{kt}]$ e $\theta_t = [\alpha_{1t} \ \alpha_{2t} \ \dots \ \alpha_{kt}]^T$ podemos escrever este modelo sob a forma de espaço de estados:

$$y_t = F_t \theta_t + \epsilon_t \quad (3.11)$$

$$\theta_t = \theta_{t-1} + \eta_t \quad (3.12)$$

ou definido o vector de caracterização $M_t = \{F_t, H_t, G_t, R_t, Q_t\}$ onde,

$$F_t = [X_{1t} \ X_{2t} \ \dots \ X_{kt}]$$

H_t matriz diagonal das variâncias covariâncias das perturbações ϵ_t
associadas às observações.

$G_t = R_t = I_k$ matriz identidade de ordem k

Q_t matriz diagonal das variâncias covariâncias das perturbações η_t
associadas aos estados.

Pode considerar-se o caso particular do modelo clássico de regressão múltipla com termo

independente, bastando para isso fazer $X_{1t}=1$ para todo t e admitir que os elementos de η_t têm variância nula, isto é os parâmetros do modelo são constantes ou seja $\theta_t = \theta_{t-1}$, para todo $t=1,2,\dots$

(2)– Modelo ARMA(p,q)

Como sabemos a equação que traduz a evolução das observações y_t segundo um processo misto autoregressivo e de média móvel de ordens respectivamente p e q é do tipo :

$$y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_p y_{t-p+1} + \eta_t + b_1 \eta_{t-1} + \dots + b_q \eta_{t-q} \quad (3.13)$$

Segundo Harvey e Philips (1979), para obter a representação de (3.13) na forma de espaço de estados, é conveniente definir ϕ_{1t} , ou b_{jt} , adicionais como apropriado (dependendo da relação de grandeza entre p e q) e reescrever (3.13) :

$$y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_r y_{t-r+1} + \eta_t + b_1 \eta_{t-1} + \dots + b_{r-1} \eta_{t-r+1} \quad (3.14)$$

onde $r = \max(p, q+1)$.

Uma representação em espaço de estados de (3.14) pode formular-se definindo um vector de estado θ_t de dimensão $(rx1)$, obedecendo à equação de transição :

$$\theta_{t+1} = G_t \theta_t + R_t \eta_t \quad \text{onde :}$$

$$G_t = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \phi_{r-1} & 0 & 0 & \dots & 1 \\ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$R_t = \begin{bmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_{r-1} \end{bmatrix}$$

e o primeiro elemento de θ_t é y_t . A equação das observações associada a (3.14) é $y_t = F_t' \theta_t$, com $F_t = [1 \ 0 \ \dots \ 0]$ vector linha de dimensão r .

A representação contemporânea e o espaço de estados são:

Com efeito da equação de transição pode escrever-se :

$$\theta_1(t+1) = \phi_1 y_t + \theta_2(t) + \eta_t \quad (1)$$

$$\theta_2(t+1) = \phi_2 y_t + \theta_3(t) + b_1 \eta_t$$

$$\vdots$$

$$\theta_{r-1}(t+1) = \phi_{r-1} y_t + \theta_r(t) + b_{r-2} \eta_t$$

$$\theta_r(t+1) = \phi_r y_t + b_{r-1} \eta_t$$

Como $\theta_1(t+1) = y_{t+1}$, substituindo $\theta_2(t)$ em (1) obtém-se ,

$$y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \theta_3(t-1) + \eta_t + b_1 \eta_{t-1}$$

(1) – Modelo Crescimento Linear e CLM

substituindo agora $\theta_3(t-1)$ vem,

Consequência a tendência é exponencial

$$y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} + \theta_4(t-2) + \eta_t + b_1 \eta_{t-1} + b_2 \eta_{t-2}$$

(2) – Modelo Crescimento Quadrático e CLM

substituindo $\theta_4(t-2)$ e assim sucessivamente obter-se-ia ,

segue-se 3. Substituição recursiva (3.15)

$$y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_r y_{t-r+1} + \eta_t + b_1 \eta_{t-1} + \dots + b_{r-1} \eta_{t-r+1}$$

Segue-se a representação com espaço de estados (3.16)

como se pretendia provar.

Segue-se 4. Representação com espaço de estados

A matriz das variâncias covariâncias das perturbações associadas às observações é nula e a matriz das variâncias covariâncias das perturbações associadas ao estado é $Q = \sigma_\eta^2 R_t R_t^T$.

segue

(3) – Modelo de Crescimento Linear.

Vimos que quando a tendência apresenta um crescimento linear se tem :

$$y_t = \mu_t + \epsilon_t \quad t=1,2,\dots,T$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + u_t$$

$$\beta_t = \beta_{t-1} + \delta_t$$

A representação correspondente em espaço de estados é:

$$y_t = F_t \theta_t + \epsilon_t$$

$$\theta_t = G_{t-1} \theta_{t-1} + \eta_t$$

com

$$F_t = [1 \ 0]; \quad \theta_t = [\mu_t \ \beta_t]^T; \quad G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ e } \eta_t = [u_t \ \delta_t]^T$$

(4) – Modelo Crescimento Linear + Ciclo

Como vimos o modelo é expresso por :

$$y_t = \mu_t + \varphi_t + \epsilon_t \quad ; \quad t = 1, 2, \dots, T$$

onde φ_t é definido por (2.34)

Assim a representação em espaço de estados é :

$$y_t = [1 \ 0 \ 1 \ 0] \theta_t + \epsilon_t \quad ; \quad \theta_t = [\mu_t \ \beta_t \ \varphi_t \ \varphi_t^*]^T$$

$$\theta_t = G_{t-1} \theta_{t-1} + \eta_t$$

com

$$G_t = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho_c \cos \lambda_c & \rho_c \sin \lambda_c \\ 0 & 0 & -\rho_c \sin \lambda_c & \rho_c \cos \lambda_c \end{bmatrix}$$

$$\eta_t = \begin{bmatrix} u_t \\ \delta_t \\ k_t \\ k_t^* \end{bmatrix}$$

A matriz de covariâncias do vector das perturbações, Q_t , é uma matriz diagonal de elementos $\{\sigma_u^2; \sigma_\delta^2; \sigma_k^2; \sigma_{k^*}^2\}$.

(5) – Modelo Estrutural Básico.

Como vimos no capítulo 2 o MEB é definido pela equação (2.2)

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

Considerando para a tendência o modelo de crescimento linear definido por (2.3) e para a sazonalidade o modelo descrito por (2.20) – sazonalidade sob a forma de factores, i.e, temos o seguinte modelo :

$$y_t = \mu_t + \gamma_t + \epsilon_t \quad (3.15)$$

$$\mu_t = \mu_{t-1} + \beta_t + u_t \quad (3.16)$$

$$\beta_t = \beta_{t-1} + \delta_t \quad (3.17)$$

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad (3.18)$$

Harvey (1989) apresenta uma formulação para o caso particular de $s=4$, generalizando para qualquer s , teremos o vector de caracterização expresso por:

$$M_t = \{ F_t, H_t, G_t, R_t, Q_t \}$$

$$F_t = [1 \ 0 \ 1 \ 0 \ \dots \ 0] ; \quad F_t \in R^{(s+1)} ; R_t = I_{(s+1)} ; H_t = [\sigma_\epsilon^2]$$

Sendo o vector de estado expresso por $\theta_t = [\mu_t \ \beta_t \ \gamma_t \ \gamma_{t-1} \ \dots \ \gamma_{t-s+2}]^T$

$$G_t = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$
$$Q_t = \begin{bmatrix} \sigma_u^2 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\delta^2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \sigma_\omega^2 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Para a sazonalidade sob a forma de Harm3nicas o modelo 3 como o anterior sendo neste caso γ_t definida por (2.22) e (2.23).

A representa33o em espa3o de estados ser3:

Equa33o das observa33es : $y_t = F_t \theta_t + e_t$

$F_t = [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ \dots \ 1 \ 0 \ 1]$; $F_t \in \mathbb{R}^{(s+1)}$

$\theta_t = \left[\mu_t \ \beta_t \ \gamma_{1,t} \ \gamma_{1,t}^* \ \dots \ \gamma_{(\frac{s}{2}-1),t} \ \gamma_{(\frac{s}{2}-1),t}^* \ \gamma_{\frac{s}{2},t} \ \gamma_{\frac{s}{2},t}^* \right]^T$ $\theta_t \in \mathbb{R}^{(s+1)}$

Equa33o de transi33o : $\theta_t = G_{t-1} \theta_{t-1} + \eta_t$; G_t 3 de ordem (s+1)

4. FILTRO DE KALMAN

4.1 Introdução

Vimos no capítulo anterior que a formulação em espaço de estados de um modelo linear (equações (3.1) e (3.3)) é bastante geral, no sentido de que qualquer modelo linear pode ser escrito como um **MLD**. O principal objectivo desta representação é o de permitir o uso do algoritmo de Kalman para fornecer previsões dos valores futuros de \mathbf{y}_t a partir do conhecimento da série observada. Pela natureza do **MLD** é óbvio que para obter tais extrapolações de \mathbf{y}_t é necessário dispor de um esquema que produza estimadores actualizados do vector de estado, θ_t . O filtro de Kalman (FK) é um conjunto de equações que nos permite estimar e corrigir as estimativas do vector de estado sempre que se dispõe de uma nova informação. Num primeiro passo, dada toda a informação disponível até ao momento presente obtém-se uma estimativa "ótima" (no sentido que minimiza o erro quadrático médio (EQM)) do estado do sistema, para o instante seguinte. Assim se dispusermos das observações $\mathbf{Y}_t = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$ estimamos o vector de estado para o instante t , isto é obtém-se a previsão a um passo de θ_t que passamos a designar por $\mathbf{a}_{t|t-1}$. Num segundo passo, quando nova observação, \mathbf{y}_t se torna disponível, esta é incorporada no algoritmo por forma a actualizar a anterior estimativa efectuada, obtendo-se $\mathbf{a}_{t|t}$. O problema da actualização da estimativa $\mathbf{a}_{t|t-1}$, isto é, a estimação de θ_t em termos das observações $\mathbf{Y}_t = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}, \mathbf{y}_t\}$ é designado por **filtragem**.

Uma outra característica do filtro de Kalman é a de permitir o **alisamento** das estimativas do vector de estado fazendo uso de observações disponíveis anterior e posteriormente ao instante t . O problema do alisamento difere assim da filtragem no sentido de que a informação acerca de θ_t não necessita estar disponível no instante t , pode haver um atraso, permitindo assim que observações obtidas após o instante t possam ser usadas na obtenção da estimativa de θ_t , que passará a designar-se por **estimativa suavizada ou alisada**, $\mathbf{a}_{t|T}$ ($T > t$) do vector de estado θ_t .

Em resumo podemos dizer que a estimação de θ_t em termos da informação

- (1) $\mathbf{Y}_{t-1} = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$ define o problema da **previsão a um passo** ($\hat{\theta}_{t|t-1} = \mathbf{a}_{t|t-1}$)
- (2) $\mathbf{Y}_t = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t\}$ define o problema da **actualização ou filtragem** ($\hat{\theta}_{t|t} = \mathbf{a}_{t|t}$)
- (3) $\mathbf{Y}_T = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t, \mathbf{y}_{t+1}, \dots, \mathbf{y}_T\}$ define o problema do **alisamento** ($\hat{\theta}_t = \mathbf{a}_{t|T}, T > t$)

Para qualquer destes problemas o **Filtro de Kalman** é o algoritmo adequado.

Vamos passar à definição do FK, abordando em primeiro lugar os problemas da previsão em um passo e da filtragem, e posteriormente os problemas da previsão em n passos e do alisamento.

4.2 DEFINIÇÃO DO FILTRO DE KALMAN

Consideremos o sistema já definido pelas equações (3.1) e (3.3)

$$y_t = F_t \theta_t + \epsilon_t \quad (4.1)$$

$$\theta_t = G_{t-1} \theta_{t-1} + R_t \eta_t \quad (4.2)$$

onde :

θ_t (mx1) vector de estado, desconhecido

G_t (mxm) matriz de transição, conhecida.

R_t (mxq) matriz associada ao ruído aleatório da equação de transição, conhecida

η_t (qx1) ruído associado à equação de estado, desconhecido.

y_t (Nx1) vector aleatório observável.

F_t (Nxm) matriz conhecida.

ϵ_t (Nx1) ruído associado às observações, desconhecido.

Com as hipóteses já atrás definidas, mas que passamos a sintetizar

- (1) Admite-se que as perturbações ϵ_t e η_t constituem " ruído branco " :

$$E[\epsilon_t] = 0 ; E[\epsilon_t \cdot \epsilon_s^T] = H_t, \forall t \text{ e } E[\epsilon_t \cdot \epsilon_s^T] = 0 \quad t \neq s$$

$$E[\eta_t] = 0 ; E[\eta_t \cdot \eta_s^T] = Q_t, \forall t \text{ e } E[\eta_t \cdot \eta_s^T] = 0 \quad t \neq s$$

- (2) $\{\epsilon_t\}$ e $\{\eta_t\}$ são processos não correlacionados

$$E[\epsilon_t \cdot \eta_s^T] = 0 \quad \forall t \text{ e } s$$

- (3) O vector de estado inicial θ_0 admite-se que tem média conhecida, a_0 e matriz de covariâncias P_0 ,

também conhecida

$$\mathbf{a}_0 = E[\theta_0] \quad ; \quad \mathbf{P}_0 = E[(\theta_0 - \mathbf{a}_0)(\theta_0 - \mathbf{a}_0)^T]$$

(4) O vector de estado inicial θ_0 é independente de ϵ_t e η_t para qualquer t .

As equações definidoras do FK que nos fornecem um algoritmo para obter estimativas "óptimas" (no sentido de que minimizam o EQM) são :

$$\mathbf{a}_{t+1|t} = (\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t) \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{y}_t \quad ; \quad \text{com } \mathbf{a}_{0|-1} = \mathbf{a}_0 \quad (4.3)$$

$$\mathbf{P}_{t+1|t} = (\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t) \mathbf{P}_{t|t-1} (\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t)^T + \mathbf{R}_{t+1} \mathbf{Q}_{t+1} \mathbf{R}_{t+1}^T + \mathbf{K}_t \mathbf{H}_t \mathbf{K}_t^T \quad (4.4)$$

onde \mathbf{K}_t é a matriz **ganho do sistema** posteriormente designada por **Ganho de Kalman** [Kalman (1960)] e cuja expressão é:

$$\mathbf{K}_t = \mathbf{G}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t]^{-1} \quad (4.5)$$

Substituindo \mathbf{K}_t pela sua expressão, em (4.4) obtém-se a chamada **equação de Riccati** (discreta)

$$\mathbf{P}_{t+1|t} = \mathbf{G}_t (\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{F}_t^T [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t]^{-1} \mathbf{F}_t \mathbf{P}_{t|t-1}) \mathbf{G}_t^T + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t^T \quad (4.6)$$

O início do processo far-se-á com $\mathbf{P}_{0|-1} = \mathbf{P}_0$.

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{F}_t^T [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t]^{-1} (\mathbf{y}_t - \mathbf{F}_t \mathbf{a}_{t|t-1}) \quad (4.7)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{F}_t^T [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t]^{-1} \mathbf{F}_t \mathbf{P}_{t|t-1} \quad (4.8)$$

As equações (4.3) e (4.4) constituem as **equações de previsão a um passo** sob a forma recursiva e as equações (4.7) e (4.8) as respectivas **equações de actualização** ou de **filtragem**.

Relativamente às equações do filtro, note-se que a actualização da estimativa do vector de estado, $\mathbf{a}_{t|t}$ (eq. 4.7) é feita à custa do erro de previsão, ν_t ,

$$\nu_t = \mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1} = \mathbf{y}_t - \mathbf{F}_t \mathbf{a}_{t|t-1} = \mathbf{F}_t (\theta_t - \mathbf{a}_{t|t-1}) + \epsilon_t \quad (4.9)$$

Os erros de previsão são designados por "inovações" visto que representam o acréscimo de informação contida na observação y_t relativamente à já existente em $\{y_0, y_1, \dots, y_{t-1}\}$. De (4.7) pode concluir-se que quanto mais o vector v_t se afasta do vector nulo, maior é a correcção (actualização) na estimativa de θ_t .

4.3 DEDUÇÃO, INTERPRETAÇÃO E PROPRIEDADES DO FILTRO DE KALMAN

Na secção anterior designou-se por $\hat{a}_{t|t-1}$ o estimador "ótimo" do vector de estado no instante t , θ_t , baseado na informação $Y_{t-1} = \{y_0, y_1, \dots, y_{t-1}\}$. Importa precisar melhor o critério de estimação e estabelecer alguns resultados antes de passar à dedução das equações do FK.

Em termos genéricos o problema que se nos depara é o seguinte: dispondo de duas variáveis aleatórias, uni ou pluridimensionais, X e Y , conjuntamente distribuídas, qual o conhecimento sobre a variável X , que nos é fornecido pelo facto de se saber que $Y=y$? Como é sabido da teoria das probabilidades a resposta encontra-se na determinação da distribuição de X condicionada por Y , $f_{X|Y}(x|y)$.

Designando por \hat{x} uma estimativa do valor assumido por X quando $Y=y$, um dos critérios possíveis para avaliar da qualidade do estimador, que a origina, é o de menor erro quadrático médio (MEQM) ou de variância mínima, se o estimador for não enviesado, i.e.,

$$E[(X - \hat{x})(X - \hat{x})^T | Y=y] \leq E[(X - z)(X - z)^T | Y=y]$$

para todos os z determinados a partir de $Y=y$.

É fácil provar que \hat{x} é única e que $\hat{x} = E[X | Y=y]$. Com efeito tem-se que:

$$\begin{aligned} E[(X - z)(X - z)^T | Y=y] &= \int_{-\infty}^{+\infty} (x - z)(x - z)^T f_{X|Y}(x|y) dx = \\ &= \int_{-\infty}^{+\infty} x x^T f_{X|Y}(x|y) dx - 2z^T \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx + z^T z = \\ &= \int_{-\infty}^{+\infty} x x^T f_{X|Y}(x|y) dx + [z^T - \int_{-\infty}^{+\infty} x^T f_{X|Y}(x|y) dx] \cdot \\ &\cdot [z - \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx] - (\int_{-\infty}^{+\infty} x^T f_{X|Y}(x|y) dx) (\int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx) \end{aligned} \quad (4.10)$$

como o primeiro e último termos do segundo membro de (4.10) não dependem de z o primeiro membro será mínimo quando $z = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx = E[X|Y=y] = \hat{x}$ e consequentemente o EQM associado à estimativa \hat{x} , obtém-se :

$$E[(X - \hat{x})(X - \hat{x})^T | Y=y] = E[\bar{X}\bar{X}^T | Y=y] - E[\bar{X}|Y=y] \cdot E[\bar{X}|Y=y]$$

Note-se que \hat{x} é uma estimativa não enviesada de X no sentido de que o valor esperado condicionado, do erro de estimação é zero, i.e ,

$$E[(X - \hat{x}) | Y=y] = E[X | Y=y] - \hat{x} = 0.$$

No caso particular de X e Y terem uma distribuição conjunta Gaussiana, isto é, sendo $W = [X^T, Y^T]^T$, W tem distribuição de Gauss com média e matriz de covariâncias, respectivamente:

$$E[W] = \begin{bmatrix} E[X] \\ E[Y] \end{bmatrix} \quad e \quad \Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

Prova-se (veja-se por exemplo Anderson (1979) ou Harvey (1989)) que a distribuição de X condicionada por $Y=y$, $f_{X|Y}(x|y)$, é também de Gauss com média e matriz de covariâncias respectivamente:

$$\hat{x} = E[X | Y=y] = E[X] + \Sigma_{XY} \Sigma_{YY}^{-1} (y - E[Y]) \quad (4.11)$$

$$\Sigma_{X|Y=y} = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T \quad (4.12)$$

A expressão (4.11) mostra-nos que \hat{x} é obtido a partir de uma transformação linear de y . Designando por $\hat{X}(Y)$ o correspondente estimador de X , i.e ,

$$\hat{X}(Y) = E[X | Y] = E[X] + \Sigma_{XY} \Sigma_{YY}^{-1} (Y - E[Y]), \text{ então}$$

$$E_{X,Y} \{ [X - \hat{X}(Y)][X - \hat{X}(Y)]^T \} = E_Y \{ E_{X|Y} \{ [X - \hat{X}(Y)][X - \hat{X}(Y)]^T | Y=y \} \} =$$

$$= E_Y \{ \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T \} = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T \quad (4.13)$$

Os índices do operador valor médio indicam sob que variáveis são considerados. Assim (4.13) prova que a matriz das covariâncias (não condicional) do erro de estimação associado com a estimativa média condicional é a mesma que a matriz das covariâncias condicionada por um particular valor de $Y=y$ (não é em geral o caso na ausência de normalidade).

Face aos resultados que acabámos de apresentar vamos agora passar à dedução das equações definidoras do FK, identificando o vector de estado com a v.a. X e considerando o caso particular em que as perturbações, ϵ_t e η_t e o vector de estado θ_0 são v.a. gaussianas. Conjugando esta hipótese com as hipóteses (1) a (4) do MLD temos que θ_0 e Y_0 são v.a. com distribuição conjunta de Gauss, i.e., $W = [\theta_0^T Y_0^T]^T$ é uma v.a. gaussiana de média $[\mathbf{a}_0^T \quad \mathbf{a}_0^T F_0^T]^T$ e matriz de covariância

$$\begin{bmatrix} \mathbf{P}_0 & \mathbf{P}_0 \mathbf{F}_0^T \\ \mathbf{F}_0 \mathbf{P}_0 & \mathbf{F}_0 \mathbf{P}_0 \mathbf{F}_0^T + \mathbf{H}_0 \end{bmatrix}$$

(i) A distribuição de θ_0 condicionada por Y_0 é também de Gauss com média

$$\mathbf{a}_{0|0} = E[\theta_0 | Y_0] = \mathbf{a}_0 + \mathbf{P}_0 \mathbf{F}_0^T [\mathbf{F}_0 \mathbf{P}_0 \mathbf{F}_0^T + \mathbf{H}_0]^{-1} (\mathbf{y}_0 - \mathbf{F}_0 \mathbf{a}_0) \quad (4.14)$$

e covariância

$$\mathbf{P}_{0|0} = \mathbf{P}_0 - \mathbf{P}_0 \mathbf{F}_0^T [\mathbf{F}_0 \mathbf{P}_0 \mathbf{F}_0^T + \mathbf{H}_0]^{-1} \mathbf{F}_0 \mathbf{P}_0 \quad (4.15)$$

(ii) Como $\theta_1 = \mathbf{G}_0 \theta_0 + \mathbf{R}_1 \eta_1$ e conjuntamente com as hipóteses de não correlação feitas, segue-se que θ_1 condicionada por $Y_0 = \{\mathbf{y}_0\}$ também tem distribuição de Gauss com média e covariância, respectivamente,

$$\mathbf{a}_{1|0} = E[\theta_1 | Y_0] = \mathbf{G}_0 \mathbf{a}_{0|0} \quad \text{e} \quad \mathbf{P}_{1|0} = \mathbf{G}_0 \mathbf{P}_{0|0} \mathbf{G}_0^T + \mathbf{R}_1 \mathbf{Q}_1 \mathbf{R}_1^T \quad (4.16)$$

(iii)

Como $\mathbf{y}_1 = \mathbf{F}_1 \theta_1 + \epsilon_1$ e $\theta_1 | Y_0$ tem distribuição normal de parâmetros dados por (4.16) segue-se

que a distribuição de \mathbf{y}_1 condicionada por \mathbf{Y}_0 é também de Gauss com média e covariância , respectivamente :

$$\hat{\mathbf{y}}_{1|0} = E [\mathbf{y}_1 | \mathbf{Y}_0] = \mathbf{F}_1 \mathbf{a}_{1|0} \quad \text{e} \quad \mathbf{F}_1 \mathbf{P}_{1|0} \mathbf{F}_1^T + \mathbf{H}_1$$

$$\text{Tem-se também que} \quad E [(\boldsymbol{\theta}_1 - \mathbf{a}_{1|0}) (\mathbf{y}_1 - \hat{\mathbf{y}}_{1|0})^T | \mathbf{Y}_0] = \mathbf{P}_{1|0} \mathbf{F}_1^T \quad (4.17)$$

Com efeito , substituindo $\mathbf{y}_1 = \mathbf{F}_1 \boldsymbol{\theta}_1 + \boldsymbol{\epsilon}_1$ e $\hat{\mathbf{y}}_{1|0} = \mathbf{F}_1 \mathbf{a}_{1|0}$, vem

$$\begin{aligned} E [(\boldsymbol{\theta}_1 - \mathbf{a}_{1|0}) (\mathbf{F}_1 \boldsymbol{\theta}_1 + \boldsymbol{\epsilon}_1 - \mathbf{F}_1 \mathbf{a}_{1|0})^T | \mathbf{Y}_0] &= \\ &= E [(\boldsymbol{\theta}_1 - \mathbf{a}_{1|0}) (\boldsymbol{\theta}_1 - \mathbf{a}_{1|0})^T \mathbf{F}_1^T + \boldsymbol{\epsilon}_1^T | \mathbf{Y}_0] \\ &= \{ E [(\boldsymbol{\theta}_1 - \mathbf{a}_{1|0}) (\boldsymbol{\theta}_1 - \mathbf{a}_{1|0})^T | \mathbf{Y}_0] \} \mathbf{F}_1^T + E [(\boldsymbol{\theta}_1 - \mathbf{a}_{1|0}) \boldsymbol{\epsilon}_1^T] = \mathbf{P}_{1|0} \mathbf{F}_1^T \end{aligned}$$

pois atendendo à independência de $\boldsymbol{\theta}_1$ e $\boldsymbol{\epsilon}_1$ e $E [\boldsymbol{\epsilon}_1] = \mathbf{0} \forall t$, o último termo anula-se.

Assim a v.a. $[\boldsymbol{\theta}_1^T \quad \mathbf{y}_1^T]$ condicionada por \mathbf{Y}_0 tem média e covariância , respectivamente

$$\begin{bmatrix} \mathbf{a}_{1|0} \\ \mathbf{F}_1 \mathbf{a}_{1|0} \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} \mathbf{P}_{1|0} & \mathbf{P}_{1|0} \mathbf{F}_1^T \\ \mathbf{F}_1 \mathbf{P}_{1|0} & \mathbf{F}_1 \mathbf{P}_{1|0} \mathbf{F}_1^T + \mathbf{H}_1 \end{bmatrix}$$

(iv)

$\boldsymbol{\theta}_1$ condicionada por $\mathbf{Y}_1 = \{\mathbf{y}_0, \mathbf{y}_1\}$ também é de Gauss com média e covariância respectivamente

$$\begin{aligned} \mathbf{a}_{1|1} &= \mathbf{a}_{1|0} + \mathbf{P}_{1|0} \mathbf{F}_1^T \left[\mathbf{F}_1 \mathbf{P}_{1|0} \mathbf{F}_1^T + \mathbf{H}_1 \right]^{-1} (\mathbf{y}_1 - \mathbf{F}_1 \mathbf{a}_{1|0}) \\ \mathbf{P}_{1|1} &= \mathbf{P}_{1|0} - \mathbf{P}_{1|0} \mathbf{F}_1^T \left[\mathbf{F}_1 \mathbf{P}_{1|0} \mathbf{F}_1^T + \mathbf{H}_1 \right]^{-1} \mathbf{F}_1 \mathbf{P}_{1|0} \end{aligned}$$

(v)

Repetindo os passos (ii) a (iv) para $t=2, \dots, t$ obtém-se :

$$\begin{aligned} \mathbf{a}_{t|t} &= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{F}_t^T \left[\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t \right]^{-1} (\mathbf{y}_t - \mathbf{F}_t \mathbf{a}_{t|t-1}) \\ \mathbf{a}_{t+1|t} &= \mathbf{G}_t \mathbf{a}_{t|t} \end{aligned}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{F}_t^T \left[\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t \right]^{-1} \mathbf{F}_t \mathbf{P}_{t|t-1}$$

$$\mathbf{P}_{t+1|t} = \mathbf{G}_t \mathbf{P}_{t|t} \mathbf{G}_t^T + \mathbf{R}_{t+1} \mathbf{Q}_{t+1} \mathbf{R}_{t+1}^T$$

De (4.9) segue-se que o vector das inovações para um filtro óptimo, e no caso Gaussiano, é tal que:

$$E[\nu_t] = 0$$

$$E[\nu_t \cdot \nu_t^T] = E[\{ \mathbf{F}_t (\theta_t - \mathbf{a}_{t|t-1}) + \epsilon_t \} \{ \mathbf{F}_t (\theta_t - \mathbf{a}_{t|t-1}) + \epsilon_t \}^T] = \mathbf{Z}_t$$

$$\text{Sendo } \mathbf{Z}_t = \mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t \quad (4.18)$$

A dedução que se acabou de apresentar, embora fácil, peca por a situação particular em que é válida, isto é, apenas sob a hipótese da normalidade. Porém é possível estabelecer as equações do FK, tal como nos trabalhos de Kalman (1960) e Kalman e Bucy (1961), baseando-nos no método das projecções ortogonais em espaços de Hilbert, sem ter necessidade das hipóteses de normalidade.

O FK pode derivar-se numa perspectiva completamente diferente da que acabámos de apresentar. Restringindo-nos à classe dos filtros lineares nas observações \mathbf{y}_t , não enviesados e de variância mínima. Consideremos a particular classe de estimativas (estimadores) assim definidas:

$$\mathbf{a}_{t+1|t} = \mathbf{K}_t^* \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{y}_t \quad (4.19)$$

e

$$E[\theta_t - \mathbf{a}_{t|t-1}] = 0 \quad (4.20)$$

onde \mathbf{K}_t^* e \mathbf{K}_t são matrizes de ponderação que vamos determinar de modo a que o estimador seja não enviesado e de variância mínima.

Definindo o erro de estimação:

$$e_{t+1|t} = \theta_{t+1} - \mathbf{a}_{t+1|t} \quad (4.21)$$

Temos que

$$\begin{aligned} e_{t+1|t} &= \mathbf{G}_t \theta_t + \mathbf{R}_{t+1} \eta_{t+1} - \mathbf{K}_t^* \mathbf{a}_{t|t-1} - \mathbf{K}_t \mathbf{y}_t = \\ &= \mathbf{G}_t \theta_t + \mathbf{R}_{t+1} \eta_{t+1} - \mathbf{K}_t^* \mathbf{a}_{t|t-1} - \mathbf{K}_t [\mathbf{F}_t \theta_t + \epsilon_t] = \\ &= [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t] \theta_t - \mathbf{K}_t^* [\theta_t - e_{t|t-1}] + \mathbf{R}_{t+1} \eta_{t+1} - \mathbf{K}_t \epsilon_t = \\ &= [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t - \mathbf{K}_t^*] \theta_t + \mathbf{K}_t^* e_{t|t-1} + \mathbf{R}_{t+1} \eta_{t+1} - \mathbf{K}_t \epsilon_t \end{aligned}$$

Estabelecemos assim a seguinte relação de recorrência para os erros de estimação

$$\mathbf{e}_{t+1|t} = [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t - \mathbf{K}_t^*] \boldsymbol{\theta}_t + \mathbf{K}_t^* \mathbf{e}_{t|t-1} + \mathbf{R}_{t+1} \boldsymbol{\eta}_{t+1} - \mathbf{K}_t \boldsymbol{\epsilon}_t \quad (4.22)$$

Para que $\mathbf{a}_{t+1|t}$ seja não enviesado, i.e. para que $E[\mathbf{e}_{t+1|t}] = 0$ tem de se verificar

$$\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t - \mathbf{K}_t^* = 0, \text{ pois por hipótese } E[\mathbf{e}_{t|t-1}] = 0, E[\boldsymbol{\eta}_t] = 0 = E[\boldsymbol{\epsilon}_t] \quad \forall t$$

Assim a relação que deve existir entre as matrizes \mathbf{K}_t e \mathbf{K}_t^* para satisfazer a condição de não enviesamento é

$$\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t = \mathbf{K}_t^*$$

Falta apenas a determinação de \mathbf{K}_t de modo a que a condição de variância mínima seja satisfeita. Substituindo \mathbf{K}_t^* em (4.22) vem :

$$\mathbf{e}_{t+1|t} = [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t] \mathbf{e}_{t|t-1} + \mathbf{R}_{t+1} \boldsymbol{\eta}_{t+1} - \mathbf{K}_t \boldsymbol{\epsilon}_t$$

pelo que a matriz de covariâncias do erro será:

$$\begin{aligned} \mathbf{P}_{t+1|t} &= E[\mathbf{e}_{t+1|t} \cdot \mathbf{e}_{t+1|t}^T] = \\ &= [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t] \mathbf{P}_{t|t-1} [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t]^T + \mathbf{R}_{t+1} \mathbf{Q}_{t+1} \mathbf{R}_{t+1}^T + \mathbf{K}_t \mathbf{H}_t \mathbf{K}_t^T \end{aligned}$$

Para que a variância seja mínima é preciso minimizar a soma dos elementos da diagonal principal da matriz $\mathbf{P}_{t+1|t}$, isto é o seu traço.

Assim \mathbf{K}_t será tal que :

$$\frac{\partial}{\partial \mathbf{K}_t} [\text{tr } \mathbf{P}_{t+1|t}] = 0 \quad \text{e} \quad \frac{\partial^2}{\partial \mathbf{K}_t^2} [\text{tr } \mathbf{P}_{t+1|t}] \text{ matriz definida positiva}$$

Para qualquer matriz \mathbf{B} simétrica tem-se:

$$\frac{\partial}{\partial \mathbf{A}} [\text{tr } \mathbf{A} \mathbf{B} \mathbf{A}^T] = 2 \mathbf{A} \mathbf{B}$$

pelo que,

$$\frac{\partial}{\partial \mathbf{K}_t} [\text{tr } \mathbf{P}_{t+1|t}] = -2 [\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t] \mathbf{P}_{t|t-1} \mathbf{F}_t^T + 2 \mathbf{K}_t \mathbf{H}_t = 0$$

ou seja :

$$- \mathbf{G}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{K}_t \mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{K}_t \mathbf{H}_t = 0$$

$$\mathbf{K}_t [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t] = \mathbf{G}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T$$

$$\mathbf{K}_t = \mathbf{G}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T [\mathbf{H}_t + \mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T]^{-1} \quad (4.23)$$

$$\frac{\partial^2}{\partial \mathbf{K}_t^2} [\text{tr } \mathbf{P}_{t+1|t}] = 2 \mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + 2 \mathbf{H}_t = 2 [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t] = 2 \mathbf{Z}_t$$

sendo \mathbf{Z}_t a matriz de covariâncias do vector das inovações é simétrica e definida positiva.

Admite-se a existência de inversa na expressão (4.23). Caso não exista substituir-se-á pela pseudo-inversa.

Encontrámos assim a expressão (4.23) para \mathbf{K}_t que é precisamente a matriz designada por **Ganho de Kalman** (4.5) e consequentemente $\mathbf{a}_{t+1|t}$ definido por (4.19) coincide com o FK inicialmente definido em (4.3).

Assim o filtro de Kalman definido pelas equações (4.3) a (4.8) é, dentro da classe dos estimadores lineares, o "óptimo" no sentido de que, sendo centrado, minimiza a variância do erro de estimação. Note-se que no caso particular dos processos gaussianos, o filtro é o estimador de variância mínima e não apenas o estimador linear de variância mínima. De realçar ainda que no caso dos processos gaussianos o FK tem uma interpretação clara e simples em termos da distribuição condicional de θ_t , isto é, o filtro fornece-nos um processo de actualização total da função de densidade de θ_t condicionada por \mathbf{Y}_t , pois que sendo esta de segunda ordem, basta para a sua actualização o conhecimento da média e matriz de covariâncias, $\mathbf{a}_{t|t-1}$ e $\mathbf{P}_{t|t-1}$, respectivamente. No caso não gaussiano $\mathbf{a}_{t|t-1}$ não é o valor esperado condicionado, mas apenas o estimador linear de variância mínima.

4.4 PREVISÃO EM N PASSOS

Para o modelo (suposto gaussiano) definido por (4.1) e (4.2) pretende-se obter a estimativa $\mathbf{a}_{t+n|t}$ do vector de estado n passos à frente, i.e, para um horizonte n , portanto no instante $t+n$, usando o conjunto de observações disponíveis $\mathbf{Y}_t = \{ \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t \}$ e consequentemente obter a previsão para \mathbf{y}_{t+n} , que designaremos por $\tilde{\mathbf{y}}_{t+n|t}$.

O filtro de Kalman fornece-nos $\mathbf{a}_{t|t}$, estimativa actualizada do vector de estado no instante t , θ_t , dado \mathbf{Y}_t , assim de (4.2) e (4.1) podemos escrever, respectivamente

$$\mathbf{a}_{t+1|t} = \mathbf{G}_t \mathbf{a}_{t|t} \quad (4.24)$$

$$\tilde{\mathbf{y}}_{t+1|t} = \mathbf{F}_{t+1} \mathbf{a}_{t+1|t} \quad (4.25)$$

Obtendo-se assim as previsões a um passo. As previsões a n passos são facilmente obtidas. Aplicando sucessivamente a equação de transição podemos escrever o vector de estado no instante $t+n$ em função do vector de estado no instante $t+1$, para o qual já se conhece um estimador (4.22).

Assim para o instante $t+2$ teremos :

$$\theta_{t+2} = \mathbf{G}_{t+1} \theta_{t+1} + \mathbf{R}_{t+2} \eta_{t+2}$$

para $t+3$,

$$\begin{aligned} \theta_{t+3} &= \mathbf{G}_{t+2} \theta_{t+2} + \mathbf{R}_{t+3} \eta_{t+3} = \mathbf{G}_{t+2} (\mathbf{G}_{t+1} \theta_{t+1} + \mathbf{R}_{t+2} \eta_{t+2}) + \mathbf{R}_{t+3} \eta_{t+3} \\ &= \mathbf{G}_{t+2} \mathbf{G}_{t+1} \theta_{t+1} + \mathbf{G}_{t+2} \mathbf{R}_{t+2} \eta_{t+2} + \mathbf{R}_{t+3} \eta_{t+3} \end{aligned}$$

Então para $t+n$ teremos:

$$\theta_{t+n} = \left(\prod_{j=1}^{n-1} \mathbf{G}_{t+j} \right) \theta_{t+1} + \sum_{j=2}^{n-1} \left(\prod_{i=j}^{n-1} \mathbf{G}_{t+i} \right) \mathbf{R}_{t+j} \eta_{t+j} + \mathbf{R}_{t+n} \eta_{t+n} \quad (4.26)$$

Como os ruídos η_t se admitem de média nula e não correlacionados com as observações \mathbf{Y}_t e, para o caso particular do modelo gaussiano, a estimativa de θ_t é a média condicionada, tomando valores esperados condicionados em (4.26) obtém-se :

$$\mathbf{a}_{t+n|t} = E [\theta_{t+n} | \mathbf{Y}_t] = \left(\prod_{j=1}^{n-1} \mathbf{G}_{t+j} \right) E [\theta_{t+1} | \mathbf{Y}_t]$$

como $E[\theta_{t+1} | Y_t] = a_{t+1|t}$ substituindo obtem-se :

$$a_{t+n|t} = \left(\prod_{j=1}^{n-1} G_{t+j} \right) a_{t+1|t} = \left(\prod_{j=0}^{n-1} G_{t+j} \right) a_t \quad (4.27)$$

que é o preditor de θ_{t+n} face à informação Y_t .

Para obter a matriz de covariâncias do erro de estimação associado àquele preditor escreva-se :

$$\begin{aligned} \theta_{t+n} - a_{t+n|t} &= \left(\prod_{j=1}^{n-1} G_{t+j} \right) \theta_{t+1} + \sum_{j=2}^{n-1} \left(\prod_{i=j}^{n-1} G_{t+i} \right) R_{t+j} \eta_{t+j} + \\ &+ R_{t+n-1} \eta_{t+n-1} - \left(\prod_{j=1}^{n-1} G_{t+j} \right) a_{t+1|t} \end{aligned}$$

Assim,

$$\theta_{t+n} - a_{t+n|t} = \left(\prod_{j=1}^{n-1} G_{t+j} \right) [\theta_{t+1} - a_{t+1|t}] + \sum_{j=2}^{n-1} \left(\prod_{i=j}^{n-1} G_{t+i} \right) R_{t+j} \eta_{t+j} + R_{t+n} \eta_{t+n}$$

Pelo que

$$\begin{aligned} P_{t+n|t} &= E \left[(\theta_{t+n} - a_{t+n|t}) (\theta_{t+n} - a_{t+n|t})^T | Y_t \right] = \\ &= \left(\prod_{j=1}^{n-1} G_{t+j} \right) P_{t+1|t} \left(\prod_{j=1}^{n-1} G_{t+j} \right)^T + \sum_{j=2}^{n-1} \left(\prod_{i=j}^{n-1} G_{t+i} \right) R_{t+j} Q_{t+j} R_{t+j}^T \left(\prod_{i=j}^{n-1} G_{t+i} \right)^T + \\ &+ R_{t+n} Q_{t+n} R_{t+n}^T \quad (4.28) \end{aligned}$$

Se o modelo for invariante tem-se que $G_t = G$; $R_t = R$ e $Q_t = Q$, $\forall t$, pelo que as expressões para a previsão em n passos simplificar-se-ão, passando a ter-se :

$$a_{t+n|t} = G^{n-1} a_{t+1|t} \quad (4.29)$$

$$P_{t+n|t} = G^{n-1} P_{t+1|t} (G^{n-1})^T + \sum_{j=2}^{n-2} G^j R Q R^T (G^j)^T \quad (4.30)$$

A previsão das observações n passos à frente será :

$$\hat{\mathbf{y}}_{t+n|t} = \mathbf{F}_{t+n} \mathbf{a}_{t+n|t} \quad (4.31)$$

A correspondente matriz de covariância do erro de previsão, que designaremos por $\mathbf{Z}_{t+n|t}$, vem

$$\begin{aligned} \text{E Q M} [\hat{\mathbf{y}}_{t+n|t}] &= \text{E} [(\mathbf{y}_{t+n} - \hat{\mathbf{y}}_{t+n|t}) (\mathbf{y}_{t+n} - \hat{\mathbf{y}}_{t+n|t})^T | \mathbf{Y}_t] = \mathbf{Z}_{t+n|t} \\ &= \text{E} [\{ \mathbf{F}_{t+n} (\boldsymbol{\theta}_{t+n} - \mathbf{a}_{t+n|t}) + \boldsymbol{\epsilon}_{t+n} \} \{ \mathbf{F}_{t+n} (\boldsymbol{\theta}_{t+n} - \mathbf{a}_{t+n|t}) + \boldsymbol{\epsilon}_{t+n} \}^T | \mathbf{Y}_t] \\ &= \mathbf{F}_{t+n} \mathbf{P}_{t+n|t} \mathbf{F}_{t+n}^T + \mathbf{H}_{t+n} \quad ; \quad n=1,2,\dots \end{aligned} \quad (4.32)$$

4.5 ALISAMENTO

Até aqui vimos como através do FK podemos resolver os problemas de previsão e de filtragem ou atualização da estimativa, isto é, baseando-nos no conjunto de observações $\mathbf{Y}_{t-1} = \{ \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{t-1} \}$ ou $\mathbf{Y}_t = \{ \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t \}$, não sendo necessário que exista atraso entre o instante em que a observação \mathbf{y}_{t-1} ou \mathbf{y}_t se torna disponível para se obterem as estimativas $\mathbf{a}_{t|t-1}$ ou $\mathbf{a}_{t|t}$. Porém se for possível existir um atraso de tempo para a produção de uma estimativa de $\boldsymbol{\theta}_t$, durante o qual novas observações se tornem disponíveis é possível incorporá-las no algoritmo de Kalman de modo a produzir estimativas de $\boldsymbol{\theta}_t$, admitindo-se um atraso de N unidades de tempo essas estimativas serão $\mathbf{a}_{t|t+N} = \text{E} [\boldsymbol{\theta}_t | \mathbf{Y}_{t+N}]$ no caso gaussiano, ou mais geralmente as estimativas lineares de menor E Q M. A $\mathbf{a}_{t|t+N}$ chamar-se-á **estimativa alisada de $\boldsymbol{\theta}_t$** e o correspondente estimador é chamado **alisador**.

É natural esperar que $\mathbf{a}_{t|t+N}$ seja uma estimativa mais precisa de $\boldsymbol{\theta}_t$, uma vez que mais informação é usada na sua determinação, isto é, alisadores em geral serão mais precisos (menor matriz de covariâncias, ou seja, a matriz diferença será semi-definida positiva) que os filtros, porém é de ter em atenção o período de atraso, se é certo que quanto maior for esse período maior precisão, será de esperar do alisador, também é certo que maior será a complexidade na sua determinação.

Conhecem-se três tipos de alisamento :

- **Alisamento de ponto fixo**, que diz respeito à obtenção de estimativas alisadas para o vector de

estado θ_t para um instante fixo $t=j$, isto é, obtenção de $a_{j|j+N}$ para um j fixo e todos os valores N .

- **Alisamento com desfasamento fixo**, diz respeito ao alisamento "on-line" dos dados, havendo um atraso fixo N entre a recepção da observação e a produção da estimativa, isto é trata-se da obtenção de $a_{t-N|t}$ para todo t e N fixo.
- **Alisamento de intervalo fixo**, diz respeito ao alisamento de um conjunto de observações, isto é, a obtenção de $a_{t|T}$ para T fixo e todos os valores de t no intervalo $0 \leq t \leq T$.

Qualquer destes três algoritmos é recursivo e estreitamente ligado ao filtro de Kalman. Para o estudo dos problemas de alisamento veja-se, por exemplo, Anderson e Moore (1979, cap.7).

4.6 ESTABILIDADE DO FILTRO DE KALMAN

Como vimos um dos principais problemas na operacionalidade do FK é a sua inicialização. E um aspecto particularmente importante quanto ao comportamento do filtro é a sua **estabilidade**, isto é, a sua capacidade de diluir o efeito das condições iniciais com o evoluir do tempo. É óbvio que um modelo deve ser estável ou pelo menos assintoticamente estável.

A forma geral do FK (eqs 4.3 a 4.5) é :

$$a_{t+1|t} = (G_t - K_t F_t) a_{t|t-1} + K_t y_t ; \quad a_{0|-1} = a_0$$

$$P_{t+1|t} = (G_t - K_t F_t) P_{t|t-1} (G_t - K_t F_t)^T + R_{t+1} Q_{t+1} R_{t+1}^T + K_t H_t K_t^T ; \quad P_{0|-1} = P_0$$

$$K_t = G_t P_{t|t-1} F_t^T [F_t P_{t|t-1} F_t^T + H_t]^T^{-1}$$

Se o modelo descrito por (4.1 e 4.2) for invariante no tempo, para que o FK seja também invariante é necessário que a matriz ganho de Kalman, K_t , não dependa do instante t , isto é, seja portanto constante, ou assintoticamente constante. Assim ter-se-á a invariância ou invariância assintótica se a equação de Riccati (4.6) admite uma solução constante ou assintoticamente constante, isto é :

$$P_{t+1|t} = \bar{P} \quad \text{ou} \quad \lim_{t \rightarrow \infty} P_{t+1|t} = \bar{P} \quad \text{sendo o correspondente ganho associado}$$

$$K_t = G \bar{P} F^T [F \bar{P} F^T + H]^{-1}$$

Um filtro invariante no tempo ou assintoticamente invariante é **estável** ou **assintoticamente estável** se os valores próprios da matriz $(G - KF)$ são todos em valor absoluto menores que 1, $(|\lambda_j(G - KF)| < 1, \forall i)$, (veja-se Anderson e Moore(1979)).

As condições para as quais um MLD é invariante no tempo e estável, estão particularmente ligadas com outras propriedades do modelo designadamente a **observabilidade** e a **controlabilidade**.

A **observabilidade** está essencialmente ligada à parametrização do vector de estado θ_t . A representação de um modelo em espaço de estados deve ser "minimal" isto é o vector de estado deve ter a menor dimensão possível.

O vector de estado diz-se **observável** se pode ser determinado exactamente a partir das observações y_t, \dots, y_{t+m-1} . Para maior facilidade na introdução do conceito de observabilidade suponhamos o MLD definido por (4.1) e (4.2), mas livre de termos irregulares, i.e., $\epsilon_t = 0$ e $\eta_t = 0 \quad \forall t$.

$$y_t = F \theta_t \quad \text{e} \quad \theta_{t+1} = G \theta_t \quad \text{donde} \quad y_{t+k} = F G^k \theta_t$$

Assim o vector das observações $\mu_t = [y_t \dots y_{t+m-1}]^T$ está relacionado com o vector de estado pela equação matricial $\mu_t = T \theta_t$ onde T é a matriz

$$T = [F^T \ (FG)^T \dots \ (F G^{m-1})^T]^T.$$

Para determinar precisamente θ_t a partir do conjunto mínimo de m observações sucessivas é pois necessário que T seja não-singular e então $\theta_t = T^{-1} \mu_t$. Assim o MLD será **observável** se a característica da matriz T for igual à dimensão, m , do vector de estado.

Uma outra propriedade importante é a **controlabilidade** que diz respeito à capacidade de um estado, θ_t , de um sistema evoluir para outro estado θ_t^* , num número finito de passos. Considerando a equação de transição (4.2) e substituindo a componente aleatória por uma componente de variáveis de controlo u_t , passando a ter-se

$$\theta_{t+1} = G \theta_t + S u_t$$

O sistema diz-se **controlável** se a matriz $[S, GS, \dots, G^{m-1}S]$ tem característica igual a m .

Note-se que, se S tem característica m a condição de controlabilidade é satisfeita, pois que, se o sistema é controlável significa que o vector de controlo u_t pode ser escolhido de modo a que o estado θ_t^* seja atingido num número finito de passos. Da equação de transição tem-se que $u_{t+1} = S^{-1}(\theta_t^* - G \theta_t)$, convertendo assim este vector de controlo θ_t em θ_t^* num único período de tempo.

5. ESTIMAÇÃO

5.1 Introdução

As equações do filtro de Kalman foram deduzidas admitindo que as matrizes do sistema (F_t , H_t , G_t , R_t e Q_t), eram totalmente conhecidas. Porém nas aplicações, em geral aquelas matrizes contêm parâmetros que são desconhecidos e que portanto precisam ser estimados com base nos dados observados $Y_T = \{y_0, y_1, \dots, y_T\}$. Designe-se por ψ o vector dos parâmetros desconhecidos. A teoria clássica de estimação pelo método de máxima Verosimilhança (M.M.V.) é possível ser aplicada aqui, embora não estejamos a trabalhar com observações independentes e idênticamente distribuídas (i.i.d.).

A distribuição conjunta das observações Y_T , fazendo uso das distribuições condicionadas, pode escrever-se :

$$f(y_0, y_1, y_2, \dots, y_T) = f(y_0) f(y_1 | y_0) f(y_2 | y_1, y_0) \dots f(y_T | y_0, y_1, \dots, y_{T-1}) \quad (5.1)$$

onde $f(y_0)$ é a função densidade de probabilidade (f.d.p.) não condicionada de y_0 e $f(y_t | Y_{t-1})$ é a densidade de probabilidade da observação y_t condicionada pelas observações $Y_{t-1} = \{y_0, y_1, \dots, y_{t-1}\}$ ($t=1, 2, \dots, T$).

Como se sabe a distribuição conjunta passa a interpretar-se como função de Verosimilhança quando encarada não como função das observações, mas sim como função dos parâmetros uma vez obtida a amostra, pelo que (5.1) pode interpretar-se como a função de Verosimilhança, que passamos a designar por $L(\cdot)$. Identificando a distribuição de ($y_0 | Y_{-1}$) com a distribuição de y_0 condicionada a não observações, isto é, não condicional portanto, tem-se

$$L(\psi, Y_T) = \prod_{t=0}^T f(y_t | Y_{t-1}) \quad (5.2)$$

5.2 ESTIMAÇÃO DE MÁXIMA VEROSIMILHANÇA. DECOMPOSIÇÃO NO ERRO DE PREVISÃO

Considerando o modelo descrito por (4.1) e (4.2) com as hipóteses adicionais de que as perturbações, ϵ_t e η_t , bem como o vector de estado inicial, são Gaussianos, segue-se que a distribuição conjunta das observações é Gaussiana bem como todas as distribuições condicionadas. Como vimos o estimador de

M.E.Q.M de \mathbf{y}_t dadas as observações \mathbf{Y}_{t-1} , coincide com o valor médio condicionado, i.e.,

$$\tilde{\mathbf{y}}_{t|t-1} = E[\mathbf{y}_t | \mathbf{Y}_{t-1}] = \mathbf{F}_t \mathbf{a}_{t|t-1} \quad (5.3)$$

e como o E Q M ($\tilde{\mathbf{y}}_{t|t-1}$) = var [$\mathbf{y}_t | \mathbf{Y}_{t-1}$] = \mathbf{Z}_t , segue-se que a distribuição de ($\mathbf{y}_t | \mathbf{Y}_{t-1}$) é Normal multivariada com média, $\tilde{\mathbf{y}}_{t|t-1}$, e matriz de variâncias covariâncias \mathbf{Z}_t dada por (4.18) assim :

$$f(\mathbf{y}_t | \mathbf{Y}_{t-1}) = (2\pi)^{-N/2} |\mathbf{Z}_t|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1})^T \mathbf{Z}_t^{-1} (\mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1})\right] \quad (5.4)$$

pois que para cada instante t , \mathbf{y}_t é um vector ($N \times 1$).

A distribuição não condicional de \mathbf{y}_0 é igualmente Normal com média $\tilde{\mathbf{y}}_0 = \mathbf{F}_0 \mathbf{a}_0$ e matriz de variâncias covariâncias $\mathbf{Z}_0 = \mathbf{F}_0 \mathbf{P}_0 \mathbf{F}_0^T + \mathbf{H}_0$, assim podemos escrever (5.2) :

$$L(\boldsymbol{\psi}, \mathbf{Y}_T) = \prod_{t=0}^T (2\pi)^{-N/2} |\mathbf{Z}_t|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1})^T \mathbf{Z}_t^{-1} (\mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1})\right] \quad (5.5)$$

usando o vector das inovações, $\boldsymbol{\nu}_t$, definido por (4.9), com $\boldsymbol{\nu}_0 = \mathbf{y}_0 - \tilde{\mathbf{y}}_0 = \mathbf{y}_0 - \mathbf{F}_0 \mathbf{a}_0$ e logaritmando vem :

$$\ln L(\boldsymbol{\psi}, \mathbf{Y}_T) = -\frac{N(T+1)}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln |\mathbf{Z}_t| - \frac{1}{2} \sum_{t=0}^T \boldsymbol{\nu}_t^T \mathbf{Z}_t^{-1} \boldsymbol{\nu}_t \quad (5.6)$$

A expressão (5.6) constitui assim a decomposição do ln da função de Verosimilhança nos erros de previsão, $\boldsymbol{\nu}_t$. Se existir informação disponível sobre todos os elementos de $\boldsymbol{\theta}_0$, o filtro de Kalman, fornecendo em cada instante t os valores, $\mathbf{a}_{t|t}$ e $\mathbf{P}_{t|t}$, permite o cálculo da função de Verosimilhança das observações \mathbf{y}_t . O problema da sua maximização relativamente aos parâmetros desconhecidos, pode ser resolvido recorrendo por exemplo a métodos numéricos de optimização não-lineares (veja-se por exemplo Harvey (1981_a)). Acontece porém que na maior parte das aplicações o valor médio, \mathbf{a}_0 , bem como a matriz de covariâncias \mathbf{P}_0 , são desconhecidos, pelo que o cálculo de (5.6) fica assim dependente do processo de inicialização do filtro de Kalman. Apresentam-se a seguir algumas soluções para este problema, destacando-se as apresentadas por Rosenberg (1973) e Jong (1988). Rosenberg propõe uma solução para o caso em que o vector de estado inicial é considerado fixo, mas desconhecido. A solução de Jong (1988) ($\mathbf{a}_0 \in \mathbf{P}_0$), mais geral, contém a de Rosenberg (1973) como um caso particular, como se verá adiante.

Quando a distribuição "à priori" de $\boldsymbol{\theta}_0$ é difusa um dos processos mais simples é inicializar o FK no instante $t = 0$ com $\boldsymbol{\theta}_0 = \mathbf{0}$ e $\mathbf{P} = \kappa \mathbf{I}$ onde κ é uma constante de valor bastante grande, mas finito,

veja-se, por exemplo, Schewppe (1973), e Harvey e Philips (1979). Após t iterações do filtro obtém-se valores aproximados para a filtragem de θ_t e respectiva matriz de covariâncias, i.e, $a_{t|t}$ e $P_{t|t}$.

Outros processos de evitar a aproximação pelo uso de " κ grande " são descritos em Ansley e Kohn(1985) que propõem uma transformação das observações de modo a eliminar a dependência das condições iniciais não especificadas.

Um processo alternativo de tratar o problema da inicialização em modelos não estacionários consiste em considerar-se um vector de estado inicial, θ_0 , constante, mas desconhecido, i.e com distribuição degenerada, $a_0 = E[\theta_0] = \theta_0$ e $P_0 = 0$. Como θ_0 é desconhecido, os seus elementos devem ser estimados, sendo pois tratados como parâmetros extra do modelo. A estimação de a_0 podia ser feita tratando os elementos de θ_0 da mesma maneira que os restantes parâmetros ψ , do modelo, e maximizando a função de Verosimilhança não-linearmente com respeito ao conjunto global de parâmetros $[\theta_0, \psi]$, porém tratando θ_0 desta maneira o processo de optimização torna-se consideravelmente complexo e pode trazer problemas de convergência devido aos erros de arredondamento. Rosenberg (1973) mostrou que se $P_0 = 0$ então o estimador de M. V. de θ_0 pode ser obtido condicionado aos outros parâmetros do modelo e eliminado da função de Verosimilhança concentrando-se esta.

O algoritmo de Rosenberg consiste em construir um FK no qual os valores iniciais são $a_0^* = 0$ e $P_0^* = 0$ produzindo um conjunto de vectores de erros de previsão (inovações), ν_0^*, \dots, ν_T^* . Para um dado valor de θ_0 o FK inicializado com $a_0 = \theta_0$ e $P_0 = 0$ produziria uma série de inovações ν_0, \dots, ν_T , que é a necessária para calcular a função de Verosimilhança (5.6). Atendendo a que as matrizes $P_{t|t-1}$, são as mesmas em ambos os filtros, pois $P_0 = 0$ em qualquer dos casos, Rosenberg vai obter as inovações ν_t a partir do "output" do 1º filtro, como $a_0^* = 0$ o vector de estado inicial pode escrever-se:

$$\theta_0 = a_0^* + \theta_0 = a_0 \quad (5.7)$$

De (4.16) temos que:

$$a_{1|0} = E[\theta_1 | Y_0] = G_0 a_{0|0} = G_0 a_0 = G_0 a_0^* + G_0 \theta_0 \quad (5.8)$$

Da equação de transição, $\theta_{t+1} = G_t \theta_t + R_{t+1} \eta_{t+1}$, decorre que para o FK inicializado com $\theta_0 = 0$ e $P_0 = 0$ se tem

$$a_{t+1|t}^* = G_t a_{t|t}^*$$

e a correspondente equação recursiva de previsão (4.3) será:

$$\mathbf{a}_{t+1|t}^* = (\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t) \mathbf{a}_{t|t-1}^* + \mathbf{K}_t \mathbf{y}_t$$

Para o FK inicializado com $\theta_0 = \mathbf{a}_0$ e $\mathbf{P}_0 = 0$ a equação (5.8) pode escrever-se :

$$\mathbf{a}_{1|0} = \mathbf{a}_{1|0}^* + \mathbf{G}_0 \theta_0 \quad (5.9)$$

para $t=1$ a equação recursiva (4.3) vem :

$$\begin{aligned} \mathbf{a}_{2|1} &= (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{a}_{1|0} + \mathbf{K}_1 \mathbf{y}_1 = (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{a}_{1|0}^* + \mathbf{K}_1 \mathbf{y}_1 + (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0 \theta_0 \\ &= \mathbf{a}_{2|1}^* + (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0 \theta_0 \end{aligned} \quad (5.10)$$

para $t=2$ vem

$$\begin{aligned} \mathbf{a}_{3|2} &= (\mathbf{G}_2 - \mathbf{K}_2 \mathbf{F}_2) \mathbf{a}_{2|1} + \mathbf{K}_2 \mathbf{y}_2 \\ &= (\mathbf{G}_2 - \mathbf{K}_2 \mathbf{F}_2) [\mathbf{a}_{2|1}^* + (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0 \theta_0] + \mathbf{K}_2 \mathbf{y}_2 \\ &= (\mathbf{G}_2 - \mathbf{K}_2 \mathbf{F}_2) \mathbf{a}_{2|1}^* + \mathbf{K}_2 \mathbf{y}_2 + [(\mathbf{G}_2 - \mathbf{K}_2 \mathbf{F}_2) (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0] \theta_0 \\ &= \mathbf{a}_{3|2}^* + [(\mathbf{G}_2 - \mathbf{K}_2 \mathbf{F}_2) (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0] \theta_0 \end{aligned} \quad (5.11)$$

para $t=t$ ter-se-á :

$$\mathbf{a}_{t+1|t} = \mathbf{a}_{t+1|t}^* + [(\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t) (\mathbf{G}_{t-1} - \mathbf{K}_{t-1} \mathbf{F}_{t-1}) \dots (\mathbf{G}_1 - \mathbf{K}_1 \mathbf{F}_1) \mathbf{G}_0] \theta_0 \quad (5.12)$$

ou definindo $\mathbf{B}_{t+1} = (\mathbf{G}_t - \mathbf{K}_t \mathbf{F}_t) \mathbf{B}_t$ com $\mathbf{B}_0 = \mathbf{G}_0$, podemos escrever (5.12) sob a forma

$$\mathbf{a}_{t+1|t} = \mathbf{a}_{t+1|t}^* + \mathbf{B}_{t+1} \theta_0 \quad (5.13)$$

Assim a previsão do vector de estado no instante $t+1$ face às observações \mathbf{Y}_t pode obter-se a partir da previsão para o FK inicializado com $\theta_0 = 0$ e $\mathbf{P}_0 = 0$, análogamente o vector dos erros de previsão, ν_t , pode decompor-se :

$$\nu_t = \mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1} = \mathbf{y}_t - \mathbf{F}_t \mathbf{a}_{t|t-1} = \mathbf{y}_t - \mathbf{F}_t (\mathbf{a}_{t|t-1}^* + \mathbf{B}_t \theta_0)$$

$$\nu_t = y_t - F_t a_{t|t-1}^* + F_t B_t \theta_0 = \nu_t^* - F_t B_t \theta_0$$

$$\nu_t = \nu_t^* - F_t B_t \theta_0 \quad (5.14)$$

Substituindo em (5.6) obtém-se :

$$\begin{aligned} \ln L(\psi, Y_T) = & - \frac{N(T+1)}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln |Z_t| - \\ & - \frac{1}{2} \sum_{t=0}^T (\nu_t^* - F_t B_t \theta_0)^T Z_t^{-1} (\nu_t^* - F_t B_t \theta_0) \end{aligned} \quad (5.15)$$

$$\begin{aligned} \ln L(\psi, Y_T) = & - \frac{N(T+1)}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln |Z_t| - \frac{1}{2} \sum_{t=0}^T \nu_t^{*T} Z_t^{-1} \nu_t^* + \\ & + \sum_{t=0}^T \theta_0^T B_t^T F_t^T Z_t^{-1} \nu_t^* - \frac{1}{2} \sum_{t=0}^T (F_t B_t \theta_0)^T Z_t^{-1} (F_t B_t \theta_0) \end{aligned} \quad (5.16)$$

Derivando (5.16) em ordem a θ_0 e igualando ao vector nulo obtém-se o estimador de M. V. para θ_0 condicionado aos outros parâmetros do modelo :

$$\hat{\theta}_0 = \left[\sum_{t=0}^T B_t^T F_t^T Z_t^{-1} F_t B_t \right]^{-1} \sum_{t=0}^T B_t^T F_t^T Z_t^{-1} \nu_t^* \quad (5.17)$$

Admite-se a existência de inversa. Se o modelo for invariante e observável a inversa existe, veja-se Harvey (1989).

Substituindo em (5.15) θ_0 por $\hat{\theta}_0$ obtém-se a função de Verosimilhança concentrada:

$$\begin{aligned} \ln L_c(\psi, Y_T) = & - \frac{N(T+1)}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln |Z_t| - \\ & - \frac{1}{2} \sum_{t=0}^T (\nu_t^* - F_t B_t \hat{\theta}_0)^T Z_t^{-1} (\nu_t^* - F_t B_t \hat{\theta}_0) = \end{aligned}$$

$$= - \frac{N(T+1)}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln |\mathbf{Z}_t| - \frac{1}{2} \sum_{t=0}^T \nu_t^* \mathbf{Z}_t^{-1} \nu_t^* + \frac{1}{2} \hat{\theta}_0^T \sum_{t=0}^T \mathbf{B}_t^T \mathbf{F}_t^T \mathbf{Z}_t^{-1} \nu_t^* \quad (5.18)$$

Note-se que :

$$\frac{1}{2} \sum_{t=0}^T \nu_t^* \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \hat{\theta}_0 - \frac{1}{2} \sum_{t=0}^T (\mathbf{F}_t \mathbf{B}_t \hat{\theta}_0)^T \mathbf{Z}_t^{-1} (\mathbf{F}_t \mathbf{B}_t \hat{\theta}_0) = 0$$

Com efeito substituindo o valor de $\hat{\theta}_0$ na 2ª parcela vem :

$$\begin{aligned} & \frac{1}{2} \sum_{t=0}^T \nu_t^* \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \hat{\theta}_0 - \frac{1}{2} \hat{\theta}_0^T \left[\sum_{t=1}^T \mathbf{B}_t^T \mathbf{F}_t^T \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \right] \times \\ & \times \left[\sum_{t=1}^T \mathbf{B}_t^T \mathbf{F}_t^T \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \right]^{-1} \sum_{t=0}^T \mathbf{B}_t^T \mathbf{F}_t^T \mathbf{Z}_t^{-1} \nu_t^* = \\ & = \frac{1}{2} \sum_{t=0}^T \nu_t^* \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \hat{\theta}_0 - \frac{1}{2} \sum_{t=1}^T \hat{\theta}_0^T \mathbf{B}_t^T \mathbf{F}_t^T \mathbf{Z}_t^{-1} \nu_t^* = 0 \end{aligned}$$

Jong(1988) deduz uma expressão para a função de Verosimilhança para o caso geral em que o vector de estado inicial θ_0 tem média \mathbf{a}_0 e matriz de covariâncias \mathbf{P}_0 não-singular, desconhecidas. O cálculo dessa função de verosimilhança também recorre ao FK inicializado com um vector de estado invariávelmente de média zero e matriz de covariâncias nula.

Observando que a distribuição conjunta de $\mathbf{Y}_T = \{ \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_T \}$ e θ_0 , (recorrendo ao teorema da probabilidade composta $P(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n) = P(\mathbf{A}_1) P(\mathbf{A}_2 | \mathbf{A}_1) \dots P(\mathbf{A}_n | \mathbf{A}_1, \dots, \mathbf{A}_{n-1})$), se pode escrever :

$$f(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_T, \theta_0) = f(\theta_0) f(\mathbf{y}_0 | \theta_0) f(\mathbf{y}_1 | \mathbf{y}_0, \theta_0) \dots f(\mathbf{y}_T | \mathbf{y}_{T-1}, \theta_0) \quad (5.19)$$

ou $f(\mathbf{Y}_T, \theta_0) = f(\mathbf{Y}_T) f(\theta_0 | \mathbf{Y}_T)$, pelo que a distribuição marginal de \mathbf{Y}_T será

$$f(\mathbf{Y}_T) = f(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_T) = \frac{f(\mathbf{Y}_T, \theta_0)}{f(\theta_0 | \mathbf{Y}_T)} \quad (5.20)$$

Assim usando (5.20) conjuntamente com (5.19), obtém-se o logaritmo da função de Verosimilhança de Y_T ,

$$\ln L(Y_T) = \ln f(\theta_0) + \ln f(y_0 | \theta_0) + \ln f(y_1 | y_0, \theta_0) + \dots + \ln(y_T | Y_{T-1}, \theta_0) - \ln f(\theta_0 | Y_T) \quad (5.21)$$

O resultado obtido por Jong(1988) é baseado num refinamento do método usado por Newbold(1974) para os modelos ARMA, conjuntamente com o resultado obtido por Rosenberg(1973) de que as inovações, ν_t , são uma combinação linear do vector de estado inicial, θ_0 , (5.14).

Assim Jong(1988) obtém para a Verosimilhança(5.21), à parte o termo constante, a seguinte expressão:

$$\begin{aligned} \ln L(Y_T) = & -\frac{1}{2} \ln |P_0| - \frac{1}{2} a_0^T P_0^{-1} a_0 - \frac{1}{2} \sum_{t=0}^T \ln |Z_t| - \frac{1}{2} \sum_{t=0}^T \nu_t^{*T} Z_t^{-1} \nu_t^* - \\ & -\frac{1}{2} \ln |P_0^{-1} + S| - \frac{1}{2} (P_0^{-1} a_0 + s)^T (P_0^{-1} + S)^{-1} (P_0^{-1} a_0 + s) \end{aligned} \quad (5.22)$$

onde ν_t^* , Z_t ($t=0,1,2,\dots,T$) tem o significado definido anteriormente isto é trata-se do vector das inovações e respectiva matriz de covariâncias para o FK inicializado com $a_0 = 0$ e $P_0 = 0$ e o vector s e matriz S são calculados em paralelo com ν_t^* e Z_t como se segue:

$$s_t = s_{t-1} + B_t^T F_t^T Z_t^{-1} \nu_t^* ; S_t = S_{t-1} + B_t^T F_t^T Z_t^{-1} F_t B_t$$

sendo B_t definido como anteriormente.

Os valores iniciais para s e S são respectivamente o vector e matriz nulos; K_t é a matriz ganho de Kalman para o filtro inicializado com $a_0 = 0$ e $P_0 = 0$.

Designando por ν^* o vector de componentes ν_t^* e por X a matriz coluna cujas linhas são os blocos $F_t B_t$ e Z a matriz diagonal cujos elementos diagonais são os blocos Z_t ($t=0,1,\dots,T$) neste caso tem-se:

$$s = X^T Z^{-1} \nu^* \quad \text{e} \quad S = X^T Z^{-1} X$$

O algoritmo de Jong (1988) inclui o de Rosenberg como um caso particular, pois que a expressão por ele apresentada para o lim de (5.22) quando P_0 tende para zero é:

$$-\frac{1}{2} \sum_{t=1}^T \ln |Z_t| - \frac{1}{2} \sum_{t=0}^T \nu_t^{*T} Z_t^{-1} \nu_t^* - \frac{1}{2} a_0^T S a_0 + s^T a_0 \quad (5.23)$$

A expressão (5.23) coincide com (5.16) visto que

$$\mathbf{a}_0 = \boldsymbol{\theta}_0, \quad \mathbf{a}_0^T \mathbf{S} \mathbf{a}_0 = \mathbf{a}_0^T \mathbf{X}^T \mathbf{Z}^{-1} \mathbf{X} \mathbf{a}_0 = \sum_{t=0}^T (\mathbf{F}_t \mathbf{B}_{t-1} \boldsymbol{\theta}_0)^T \mathbf{Z}_t^{-1} (\mathbf{F}_t \mathbf{B}_t \boldsymbol{\theta}_0)$$

$$\mathbf{s}^T \mathbf{a}_0 = \sum_{t=0}^T \nu_t^T \mathbf{Z}_t^{-1} \mathbf{F}_t \mathbf{B}_t \boldsymbol{\theta}_0$$

A função de Verossimilhança (5.23) pode ser concentrada com respeito a \mathbf{a}_0 e \mathbf{P}_0 . O estimador que Jong(1988) obtém para $\mathbf{a}_0 = \boldsymbol{\theta}_0$ é $\hat{\boldsymbol{\theta}}_0 = \mathbf{S}^{-1} \mathbf{s}$ que coincide precisamente com (5.17) obtido por Rosenberg. Pode provar-se que uma condição suficiente para que \mathbf{S} seja não singular para um modelo invariante é que $\boldsymbol{\theta}_0$ seja observável.

No seu trabalho Jong (1988) apresenta ainda a expressão da função de Verossimilhança para o caso em que a distribuição "à priori" de $\boldsymbol{\theta}_0$ é difusa, expressão que pode ser obtida de (5.22) fazendo \mathbf{P}_0 tender para infinito de modo que \mathbf{P}_0^{-1} tenda para zero. Obtém assim a função de Verossimilhança baseada numa transformação linear das observações de modo a tornar os dados invariantes a $\boldsymbol{\theta}_0$ como em Ansley & Kohn(1985).

6. APLICAÇÃO DO FILTRO DE KALMAN A MODELOS UNIVARIADOS.

6.1 Introdução

Neste capítulo apresentam-se os principais resultados decorrentes da aplicação da teoria do FK para o caso particular dos modelos estruturais univariados, dando particular destaque ao modelo estrutural básico (M.E.B), que foi descrito no capítulo 2 e cuja representação sob a forma de espaço de estados se apresentou no cap.3 pags 22 a 24.

A formulação geral de um modelo univariado sob a forma de espaço de estados obtém-se de (4.1) e (4.2), fazendo $N=1$, assim y_t representa apenas uma variável, a matriz F_t passará a ser um vector transposto $(1 \times m)$, f_t , e a matriz H_t , das covariâncias das perturbações associadas às observações passa a ser um escalar, que designaremos por, h_t , assim como a matriz, Z_t , das covariâncias do vector das inovações, v_t , passará a escalar, z_t .

As equações do FK (4.3) a (4.8) são idênticas, apenas com as observações que se acabam de fazer.

6.2 - ESTIMAÇÃO DE MÁXIMA VEROSIMILHANÇA NO DOMÍNIO DO TEMPO

A expressão para a função de Verosimilhança de um modelo estrutural univariado obtém-se directamente de (5.5) ou o seu logaritmo de (5.6) notando que $N=1$ e substituindo a matriz Z_t pelo correspondente escalar z_t , assim obtém-se :

$$\ln L = - \frac{T+1}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^T \ln z_t - \frac{1}{2} \sum_{t=0}^T v_t^2 / z_t \quad (6.1)$$

O modelo estrutural básico (M.E. B) como decorre da sua representação em espaço de estados é invariante no tempo. Considerando a sazonalidade modelada sob a forma de factores (eqs 3.15 a 3.18) contém quatro parâmetros desconhecidos, $h = \sigma_\epsilon^2$, σ_u^2 , σ_δ^2 e σ_ω^2 , respectivamente as variâncias da componente irregular da equação das observações, variância da componente irregular associada à tendência, variância da componente irregular associada ao declive da tendência e variância da

componente irregular associada à sazonalidade. Contudo otimização não linear necessita apenas de ser feita com respeito a três daqueles parâmetros, visto que um deles por exemplo, σ_{ϵ}^2 , pode ser previamente estimado em função dos outros, obtendo-se o FK como uma função das variâncias relativas $\sigma_u^2 / \sigma_{\epsilon}^2$; $\sigma_{\delta}^2 / \sigma_{\epsilon}^2$ e $\sigma_{\omega}^2 / \sigma_{\epsilon}^2$.

Como vimos no capítulo anterior o problema do cálculo de (6.1) é o da inicialização do FK, isto é a distribuição "à priori" do vector de estado no instante $t=0$. É evidente que qualquer dos métodos descritos [Rosenberg(1973) e Jong(1988)] se aplicam aqui. Porém para modelos univariados, o uso de uma distribuição "à priori" difusa é equivalente à construção de uma distribuição própria a partir de um conjunto inicial de observações desde que o modelo seja observável [Harvey(1989)].

Assumindo que os K primeiros valores da série são utilizados para a construção dos valores iniciais do vector de estado $\mathbf{a}_{k|k}$ e $\mathbf{P}_{k|k}$ a função de Verosimilhança (6.1), para o caso do M.E.B ($k = s+1$, dimensão do vector de estado), fazendo $\mathbf{z}_t = \text{var}[\nu_t] = \sigma_{\epsilon}^2 \mathbf{z}_t^*$, e designando por ψ , o vector dos parâmetros do modelo, segundo Harvey & Todd (1983), pode escrever-se :

$$\ln L(\psi | y_{s+1}, \dots, y_T | Y_s) = -\frac{T-s}{2} \ln 2\pi - \frac{T-s}{2} \ln \sigma_{\epsilon}^2 - \frac{1}{2} \sum_{t=s+1}^T \ln z_t^* - \frac{1}{2\sigma_{\epsilon}^2} \sum_{t=s+1}^T \nu_t^2 / z_t^* \quad (6.2)$$

Derivando (6.2) em ordem a σ_{ϵ}^2 , como ν_t e \mathbf{z}_t^* não dependem de σ_{ϵ}^2 , obtém-se :

$$\frac{\partial \ln L}{\partial \sigma_{\epsilon}^2} = -\frac{T-s}{2} (\sigma_{\epsilon}^2)^{-1} + \frac{1}{2\sigma_{\epsilon}^4} \sum_{t=s+1}^T \nu_t^2 / z_t^* \quad (6.3)$$

Pelo que o estimador de máxima verosimilhança para σ_{ϵ}^2 , condicionado aos outros parâmetros, ψ^* , ($\psi = [\psi^*, \sigma_{\epsilon}^2]$) vem :

$$\hat{\sigma}_{\epsilon}^2(\psi^*) = \frac{1}{T-s} \sum_{t=s+1}^T \nu_t^2 / z_t^* \quad (6.4)$$

Substituindo em (6.2) obtém-se a função de verosimilhança concentrada, L_c , reduzindo-se assim em um o número de parâmetros

$$\ln L_c(\psi) = -\frac{T-s}{2} \ln(2\pi + 1) - \frac{1}{2} \sum_{t=s+1}^T \ln z_t^* - \frac{T-s}{2} \ln \hat{\sigma}_{\epsilon}^2(\psi^*) \quad (6.5)$$

esta função agora deverá ser maximizada com respeito aos restantes elementos de ψ .

Harvey e Peters (1984) apresentam um processo alternativo para a estimação de θ_0 , que consiste na sua estimação pelo método dos mínimos quadrados generalizado (GLS). Como já se disse o M.E.B é invariante no tempo, pelo que as equações que o traduzem sob a forma de espaço de estados podem escrever-se:

$$y_t = f \theta_t + \epsilon_t \quad (6.6)$$

$$\theta_t = G \theta_{t-1} + \eta_t \quad (6.7)$$

y_t pode exprimir-se em termos de θ_0 , por substituições sucessivas vem:

$$\begin{aligned} y_t &= f (G \theta_{t-1} + \eta_t) + \epsilon_t = \\ &= f G \theta_{t-1} + f \eta_t + \epsilon_t = \\ &= f G (G \theta_{t-2} + \eta_{t-1}) + f \eta_t + \epsilon_t = \\ &= f G^2 \theta_{t-2} + f G \eta_{t-1} + f \eta_t + \epsilon_t = \\ &= f G^2 (G \theta_{t-3} + \eta_{t-2}) + f G \eta_{t-1} + f \eta_t + \epsilon_t = \\ &= f G^3 \theta_{t-3} + f G^2 \eta_{t-2} + f G \eta_{t-1} + f \eta_t + \epsilon_t = \\ &= \dots = f G^t \theta_0 + f G^{t-1} \eta_1 + \dots + f G^2 \eta_{t-2} + f G \eta_{t-1} + f \eta_t + \epsilon_t \end{aligned}$$

Assim podemos escrever:

$$y_t = f G^t \theta_0 + f \sum_{j=0}^{t-1} G^j \eta_{t-j} \quad (6.8)$$

Definindo

$$\xi_t = f \sum_{j=0}^{t-1} G^j \eta_{t-j} \quad (6.9)$$

$$X_t = f G^t \quad (6.10)$$

Podemos escrever (6.8) sob a forma matricial

$$y = X \theta_0 + \xi \quad (6.11)$$

Designando a matriz das covariâncias de ξ_t por Ω_0 , $\text{var} [\xi \xi^T] = \Omega_0$, o estimador GLS de θ_0 é como se sabe

$$\hat{\theta}_0 = [X^T \Omega_0^{-1} X]^{-1} X^T \Omega_0^{-1} y \quad (6.12)$$

A função de Verosimilhança concentrada será:

$$\ln L_c(\psi, Y_T) = -\frac{(T+1)}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega_0| - \frac{1}{2} (y - X \hat{\theta}_0)^T \Omega_0^{-1} (y - X \hat{\theta}_0) \quad (6.13)$$

Segundo Harvey e Peters(1984, apêndice B) as expressões da função de verosimilhança (6.13) e (6.1) são equivalentes.

6.3 ESTIMAÇÃO A PARTIR DA FORMA REDUZIDA. CORRELOGRAMA

Um modelo estrutural para séries cronológicas, em geral, contém vários termos residuais, ou perturbações. Desde que o modelo seja linear, as diferentes componentes podem ser combinadas de forma a conduzir a um modelo com uma única componente aleatória, ou seja, à **forma reduzida ou canónica do modelo**. A forma reduzida é um modelo ARIMA com várias restrições no espaço dos parâmetros. Se estas restrições não são impostas quando o modelo ARIMA é ajustado, diz-se que se está a tratar da **forma reduzida não restrita**. A forma reduzida para o modelo estrutural básico foi apresentada em (2.29) e é tal que $\Delta \Delta_s y_t$ segue um processo de médias móveis de ordem $(s+1)$, $MA(s+1)$, onde os parâmetros estão sujeitos a um certo número de restrições.

A forma reduzida não restrita contém em geral mais parâmetros do que a forma estrutural, assim para o MEB, teremos $(s+2)$ em vez de 4.

A estimação via forma reduzida não restrita, pelos métodos Box-Jenkins tradicionais, pode conduzir a que os estimadores obtidos para os parâmetros estruturais não sejam admissíveis. Porém, um problema mais grave está relacionado com o facto, que já foi mencionado, de que a correspondência entre os parâmetros estruturais e os da forma reduzida, não é de um para um, o que conduz a um problema de sobreidentificação, sendo possível encontrar-se diferentes soluções sem que a priori seja possível dizer qual delas é a melhor.

A estimação da forma reduzida através da maximização da função de verosimilhança com respeito aos parâmetros estruturais, garante a satisfação das restrições de não negatividade das variâncias bem

como todas as outras, mas o método geral de cálculo desenvolvido por Nerlove e al (1979) é bastante complexo.

Parece não haver vantagens em estimar os parâmetros do modelo estrutural via forma reduzida, pois se se pretende estimadores exactos de M.V., tão complicado e moroso é obtê-los para um modelo ARIMA como para um modelo estrutural. Porém, os estimadores obtidos a partir do correlograma podem ser úteis como estimadores preliminares ou como pontos de partida para processos iterativos, usados na maximização da função de verosimilhança.

Como vimos para o Modelo Estrutural Básico a sua forma estacionária (2.29), é

$$\Delta \Delta_s y_t = S(L) \delta_{t-1} + \Delta_s u_t + \Delta^2 \omega_t + \Delta \Delta_s \epsilon_t = \\ = \sum_{j=1}^s \delta_{t-j} + u_t - u_{t-s} + \omega_t - 2\omega_{t-1} + \omega_{t-2} + \epsilon_t - \epsilon_{t-1} - \epsilon_{t-s} + \epsilon_{t-s-1}; t=s+2, \dots$$

Partindo desta expressão é fácil calcular a função de autocovariância do processo estacionário $Z_t = \Delta \Delta_s y_t$. Como $E[Z_t] = 0$, a função de autocovariância reduz-se a $\gamma(\tau) = E[Z_t \cdot Z_{t-\tau}]$ e obtém-se

$$\begin{aligned} \gamma(0) &= s\sigma_\delta^2 + 2\sigma_u^2 + 4\sigma_\omega^2 + 4\sigma_\epsilon^2 \\ \gamma(1) &= (s-1)\sigma_\delta^2 - 4\sigma_\omega^2 - 2\sigma_\epsilon^2 \\ \gamma(2) &= (s-2)\sigma_\delta^2 + \sigma_\omega^2 \\ &\dots \\ \gamma(\tau) &= (s-\tau)\sigma_\delta^2 \quad ; \tau = 3, \dots, s-2 \\ &\dots \\ \gamma(s-1) &= \sigma_\delta^2 + \sigma_\epsilon^2 \\ \gamma(s) &= -\sigma_\eta^2 - \sigma_\epsilon^2 \\ \gamma(s+1) &= \sigma_\epsilon^2 \\ \gamma(\tau) &= 0 \quad ; \tau \geq s+2 \end{aligned} \quad (6.14)$$

Assumindo que todas as variâncias no modelo são estritamente positivas, as equações em (6.14) implicam as seguintes restrições nas autocorrelações ($\rho(\tau) = \gamma(\tau)/\gamma(0)$):

$$\begin{aligned} \rho(\tau)/\rho(\tau+1) &= (s-\tau)/(s-1-\tau) \quad ; \tau = 3, \dots, s-3 \\ \rho(\tau) &> 0 \\ \rho(s) &< 0 \\ \rho(\tau-1) &> \rho(\tau+1) > 0 \\ |\rho(s)| &> \rho(s+1) \end{aligned} \quad (6.15)$$

Segundo Nerlove e al (1979) estimadores aproximados de M.V. para os parâmetros estruturais podem ser mais facilmente calculados no domínio da frequência.

6.4 ESTIMÇÃO DO M.E.B. NO DOMÍNIO DA FREQUÊNCIA

Considerando a forma estacionária do M.E.B. , (2.29) , e $T^* = T \cdot (s+1)$, pode provar-se, Harvey e Peters (1984) , Harvey (1989), que a função de verosimilhança para Δ, Δ_s, y_t pode apresentar-se sob a forma :

$$\ln L = - T^* \ln 2\pi - \frac{1}{2} \sum_{j=0}^{T^*-1} \ln f(\lambda_j) - \frac{1}{2} \sum_{j=0}^{T^*-1} \frac{I(\lambda_j)}{f(\lambda_j)} \quad (6.16)$$

onde $I(\lambda_j)$ e $f(\lambda_j)$ representam respectivamente a função de densidade espectral amostral ou periodograma e função densidade espectral teórica de Δ, Δ_s, y_t para a frequência $\lambda_j = 2\pi j / T^*$ ($j=0,1,...,T^*-1$) . Definidas por,

$$I(\lambda_j) = \frac{1}{2\pi T^*} \left| \sum_{t=1}^{T^*} y_t e^{-i \lambda_j t} \right|^2 ; (j=0,1,...,T^*-1) \quad (6.17)$$

$$f(\lambda_j) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i \lambda_j \tau) \quad (6.18)$$

onde $i = \sqrt{-1}$ e $\gamma(\tau)$ a função de autocorrelação.

A função geradora das autocovariâncias(f.g.ac.) para um processo estacionário é definida como um polinómio no operador atraso, $g(L)$, tal que ,

$$g(L) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) L^\tau$$

assim a função de verosimilhança (6.16) pode exprimir-se em termos da função geradora das autocovariâncias, pois que , para $L = \exp(-i \lambda_j)$ tem-se a relação :

$$f(\lambda_j) = (2\pi)^{-1} g(\exp(-i \lambda_j))$$

Designando $\mathbf{g}(\exp(i\lambda_j))$ por \mathbf{g}_j e substituindo em (6.16) vem :

$$\begin{aligned} \ln L &= -T^* \ln 2\pi - \frac{1}{2} \sum_{j=0}^{T^*-1} \ln \left(\frac{1}{2\pi} \mathbf{g}_j \right) - \frac{1}{2} \sum_{j=0}^{T^*-1} \frac{I(\lambda_j)}{\left(\frac{1}{2\pi} \mathbf{g}_j \right)} \\ &= -\frac{T^*}{2} \ln 2\pi - \frac{1}{2} \sum_{j=0}^{T^*-1} \ln \mathbf{g}_j - \pi \sum_{j=0}^{T^*-1} \frac{I(\lambda_j)}{\mathbf{g}_j} \end{aligned} \quad (6.19)$$

As expressões (6.16) e (6.19) representam apenas a f.v. aproximada para $\Delta \Delta_{\mathbf{y}_1}$. Pois que , segundo Harvey e Peters (1984) e Harvey(1989) aquelas expressões apenas são iguais à função de verosimilhança se o processo for " processo circular " (i. e. se a matriz das variâncias covariâncias do processo for tal que $\gamma(\tau) = \gamma(T^* - \tau)$ $\tau = 1, \dots, T^* - 1$), condição que em geral não se verifica.

Como a f.g.ac é função dos parâmetros desconhecidos, ψ , a maximização da f.v. envolve iterações com diferentes valores de ψ e por conseguinte de \mathbf{g}_j , até que (6.18) seja maximizada. As ordenadas do periodograma são independentes de ψ e portanto serão calculadas apenas uma vez e não em cada iteração do processo numérico de optimização da função de verosimilhança.

Harvey (1989) calcula para o M. E. B a função de densidade espectral , obtendo ,

$$\mathbf{g}_j = z_{\mu j} \sigma_\eta^2 + z_{\beta j} \sigma_\delta^2 + z_{\gamma j} \sigma_\omega^2 + z_{\epsilon j} \sigma_\epsilon^2 \quad (6.20)$$

onde

$$z_{\mu j} = 2(1 - \cos \lambda_j s) \quad (6.21)$$

$$z_{\beta j} = (1 - \cos \lambda_j s) / (1 - \cos \lambda_j) = s + 2 \sum_{h=1}^{s-1} (s-h) \cos \lambda_j h \quad (6.22)$$

$$z_{\gamma j} = 6 - 8 \cos \lambda_j + 2 \cos 2 \lambda_j \quad (6.23)$$

$$\text{e } z_{\epsilon j} = 4(1 - \cos \lambda_j s) (1 - \cos \lambda_j) \quad (6.24)$$

Tal como no domínio do tempo um dos parâmetros , geralmente σ_ϵ^2 ou σ_η^2 pode ser eliminado da função de verosimilhança , suponha-se que é σ_ϵ^2 e que $\mathbf{g}_j = \sigma_\epsilon^2 \mathbf{g}_j^*$, com σ_η^2 , σ_δ^2 e σ_ω^2 substituídas pelas variâncias relativas $q_\eta = \sigma_\eta^2 / \sigma_\epsilon^2$, $q_\delta = \sigma_\delta^2 / \sigma_\epsilon^2$ e $q_\omega = \sigma_\omega^2 / \sigma_\epsilon^2$. O vector dos parâmetros , ψ_* , é

$\psi_* = [q_\eta , q_\delta , q_\omega]^T$, substituindo $\mathbf{g}_j = \sigma_\epsilon^2 \mathbf{g}_j^*$ na expressão da função de verosimilhança e maximizando em relação a σ_ϵ^2 obtém-se :

$$\hat{\sigma}_\epsilon^2 = \frac{2\pi}{T^*} \sum_{j=0}^{T^*-1} \frac{I(\lambda_j)}{\xi_j^2} \quad (6.25)$$

Nerlove e al (1979) fizeram bastante uso da estimação de modelos UCARIMA no domínio da frequência, tendo concluído que do ponto de vista computacional são bastante mais rápidos do que os processos de estimação no domínio do tempo. Contudo, eles não consideraram os processos de estimação no domínio do tempo, baseados no filtro de Kalman.

Harvey e Peters (1984) aplicaram os vários processos de estimação de máxima verosimilhança a dois conjuntos de séries, um contendo três séries trimestrais de 60 observações, simuladas para diferentes valores das variâncias no M.E.B., sendo as perturbações geradas por um processo que produz variáveis Normais i.i.d., e o outro conjunto contendo sete séries reais, também trimestrais e relacionadas com índices macroeconómicos para o Reino Unido.

Verificaram algumas diferenças em relação às estimativas obtidas pelos processos no domínio do tempo (DT) e no domínio da frequência (DF), em particular o processo DF produziu sempre estimativas positivas para a variância das perturbações do declive da tendência (σ_δ^2), e para a variância das perturbações da sazonalidade (σ_ω^2), o que se deve ao facto de que a estimação no DF não permitir detectar situações em que $\sigma_\omega^2=0$ ou $\sigma_\delta^2=0$, pois que para aqueles valores a f.v. no DF é não limitada, contrariamente ao que acontece com a função de verosimilhança no DT.

Para as séries reais, as estimativas de σ_u^2 no DF apresentam um valor superior aos da estimação no DT. Por outro lado as estimativas de σ_ϵ^2 estão por vezes muito perto de zero pelo que a concentração da função de verosimilhança será preferível fazer-se em relação a σ_u^2 ou σ_δ^2 em vez de σ_ϵ^2 .

As estimativas obtidas no DT e DF são semelhantes em particular quando no DT o FK converge rapidamente para o seu estado estacionário.

Harvey e Peters também concluíram que do ponto de vista computacional a estimação no domínio da frequência é substancialmente mais rápida que a estimação no domínio do tempo. Como se fez notar a f.v. no domínio da frequência só é possível calcular exactamente se as condições de "circularidade" forem satisfeitas, porém dada a maior rapidez, as estimativas obtidas no DF podem servir para ponto de partida no DT.

É possível provar que as propriedades assintóticas dos estimadores de M.V. obtidos a partir de um conjunto de observações i.i.d., continuam válidas quando aquelas condições se relaxam. Assim os estimadores de M.V. obtidos, quer no domínio do tempo, quer no domínio da frequência são assintoticamente consistentes e Normais.

Crowder (1976) demonstra a consistência fraca e a normalidade assintótica estabelecendo as

condições para que isso se verifique.

Hannan(1970) deduz as propriedades assintóticas dos E.M.V. no DF para modelos estruturais. Considerando um processo estacionário Gaussiano com função de densidade espectral, $f(\lambda)$, contínua, não se anulando no intervalo $[-\pi, \pi]$ e dependendo de um conjunto ψ de parâmetros desconhecidos prova que :

- (i) $\hat{\psi}$, o estimador de M. V., converge em probabilidade para ψ (plim $\hat{\psi} = \psi$) ;
- (ii) Sob as hipóteses de continuidade para as derivadas até à 3ª ordem com respeito a ψ da função inversa da f.g.ac. $1/g(L)$, na vizinhança de ψ e admitindo que os parâmetros θ_h , na representação de médias móveis (MA) do modelo

$$y_t = \sum_{h=0}^{\infty} \theta_h \zeta_{t-h} \quad ; \quad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2) \quad (6.26)$$

são tais que : $\sum_{h=0}^{\infty} h |\theta_h| < \infty$ então $\sqrt{T}(\hat{\psi} - \psi)$ tem distribuição

limite Normal com média zero e matriz de covariâncias $\mathbf{IA}^{-1}(\psi)$ com $\mathbf{IA}(\psi)$ dada por

$$\mathbf{IA}(\psi) = \frac{1}{4\pi} \int_{-\pi}^{+\pi} \frac{\partial \ln g(e^{i\lambda})}{\partial \psi} \frac{\partial \ln g(e^{i\lambda})}{\partial \psi^T} d\lambda$$

Se o processo for identificável, $\mathbf{IA}(\psi)$ é não-singular. No caso do modelo estrutural básico as condições exigidas são satisfeitas se $\sigma_\delta^2 > 0$ e $\sigma_\omega^2 > 0$.

7. MODELOS ESTRUTURAIS COM VARIÁVEIS EXPLICATIVAS E VARIÁVEIS DE INTERVENÇÃO

7.1. VARIÁVEIS EXPLICATIVAS

A variável y_t nos modelos estruturais considerados anteriormente foi modelada exclusivamente em termos dos valores do seu passado. Por exemplo no M. E.B os movimentos de y_t são explicados em termos de uma tendência estocástica e uma componente sazonal também estocástica. Porém frequentemente conhecem-se outras séries que têm uma relação de casualidade com a variável y_t em estudo. Nesse caso é possível, se as variáveis adicionais são exógenas e a relação entre elas e y_t é linear reescrever o M.E.B. :

$$y_t = \mu_t + \gamma_t + \sum_{i=1}^k \alpha_i x_{it} + \epsilon_t \quad ; t=1, \dots, T \quad (7.1)$$

onde x_{it} são as variáveis explicativas e α_i os parâmetros que lhe estão associados. Mais geralmente outras componentes como ciclos ou efeitos diários podem ser usadas , assim para o modelo estrutural completo ter-se-ia :

$$y_t = \mu_t + \gamma_t + \varphi_t + \sum_{i=1}^k \alpha_i x_{it} + \epsilon_t \quad ; t=1, \dots, T \quad (7.2)$$

É evidente que as variáveis explicativas apenas contribuem em parte para a explicação dos movimentos de y_t , pois de contrário ter-se-ia apenas um modelo de regressão linear.

A representação de (7.1) em espaço de estados pode exprimir-se de uma maneira semelhante ao que foi feito em (3.1) ,mantendo a mesma equação de transição e considerando como equação das observações :

$$y_t = f_t \theta_t + x_t^T \alpha + \epsilon_t \quad (7.3)$$

Se o vector dos parâmetros associado às variáveis exógenas for conhecido a adição de $x_t^T \alpha$ à equação das observações não traria qualquer problema à aplicação do FK . Contudo na maior parte dos casos α é

desconhecido. Um processo será incluí-lo no próprio vector de estado, passando a ter-se um vector de estado alargado $\theta_t^* = [\theta_t^T \alpha_t^T]^T$ que satisfaz ao sistema :

$$y_t = [f_t \ x_t^T] \theta_t^* + \epsilon_t \quad (7.4)$$

$$\theta_t^* = \begin{bmatrix} \theta_t \\ \alpha_t \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \alpha_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \end{bmatrix} \quad (7.5)$$

Este modelo de espaço de estados aumentado é conveniente para previsão visto que permite actualizar simultaneamente o estimador de α sempre que uma nova observação se torna disponível . A estimação pode ser feita por qualquer dos processos descritos anteriormente quer no domínio do tempo quer no domínio da frequência.

7.2 VARIÁVEIS DE INTERVENÇÃO

Variáveis de intervenção, correspondentes às variáveis artificiais ou "dummies" nos modelos de regressão, podem também ser incluídas no modelo por forma a traduzir os efeitos de determinados acontecimentos no comportamento da série em estudo. Na análise de intervenção pura não se consideram presentes outras variáveis explicativas no modelo .

O modelo (7.4) pode assim ser generalizado :

$$y_t = \mu_t + \gamma_t + \varphi_t + \sum_{i=1}^k \alpha_i x_{it} + \sum_{j=1}^k \lambda_j W_{jt} + \epsilon_t \quad (7.6)$$

onde W_{jt} representam as variáveis de intervenção e λ_j os seus coeficientes. A definição de W_{jt} depende da forma que o efeito de intervenção pode assumir. Por exemplo se o efeito da intervenção é transitório , i.e. se tem efeito apenas no instante $t=\tau$, ter-se-á :

$$W_{jt} = \begin{cases} 0 & t \neq \tau \\ 1 & t = \tau \end{cases}$$

Estimação de um modelo da forma (7.6) pode igualmente efectuar-se quer no DT quer no DF ,

tratando a variável de intervenção como se fosse mais uma variável explicativa.

A análise de intervenção foi introduzida por Box e Tiao(1975) para modelos ARIMA. Harvey e Durbin(1986) na sua análise do efeito da legislação sobre o uso do cinto de segurança em viaturas automóveis, no número de pessoas mortas e gravemente feridas , põem bem patente a importância da possibilidade de alargar os modelos estruturais à inclusão de quer variáveis explicativas quer variáveis de intervenção. Voltaremos a este assunto na segunda parte deste trabalho.

SEGUNDA PARTE

Modelos Estruturais - Abordagem Bayesiana.

8 - MODELOS ESTRUTURAIS - ABORDAGEM BAYESIANA.

8.1 - Introdução.

A metodologia bayesiana de previsão para modelos estruturais foi introduzida por Harrison e Stevens (1971/76). Os modelos bayesianos geralmente operam segundo o "Princípio de Gestão por Excepção" ("Management by Exception"). Isto é, rotinas de previsão são produzidas por um modelo estatístico e usadas como "input" para um processo de decisão, excepto se circunstâncias excepcionais se verificam. Por exemplo, o aparecimento de informação relevante proveniente de uma fonte externa ao sistema, a antecipação de uma greve laboral, nova legislação, uma grande campanha promocional, são situações que afectam a procura e aumentam a incerteza futura. Outros tipos de excepção são retrospectivos e ocorrem quando se verifica uma mudança imprevisível, por exemplo, no padrão de procura. Refira-se, como relevante, que as excepções constituem uma oportunidade para aprendizagem e melhor conhecimento do mercado.

Um sistema de previsão (modelo + utilizador) deve permitir uma análise retrospectiva ("What happened analysis") e uma análise prospectiva ("What if analysis"), devendo portanto ser um modelo robusto e estruturado apropriadamente.

A abordagem bayesiana para a modelação de um sistema de previsão deve constar essencialmente de duas fases :

- (1) descrição do estado corrente do processo
- (2) relacionamento entre os estados corrente e futuro do processo.

Na fase (1) a principal preocupação deve ser a comunicação entre o agente da decisão e o modelo estatístico. Quais os parâmetros que caracterizam o processo? Que informação disponível existe sobre o estado corrente do processo? Como descrever essa informação em termos, de "melhores" estimativas pontuais, de medidas de incerteza ou de distribuições de probabilidade?

As principais características do sistema de previsão bayesiana segundo J.Harrison e West são :

Representação paramétrica (ou em espaço de estados) - ao contrário da funcional, como acontece, por exemplo, nos modelos ARIMA - o que significa que os modelos são analisados na sua forma estrutural, atribuindo-se a cada elemento do vector dos parâmetros, uma interpretação específica;

Descrição probabilística dos parâmetros em qualquer instante especificado, exprimindo estas crenças, $f(\theta_i | D_i)$, acerca do estado corrente θ_i dada a informação disponível no instante t , D_t .

Definição sequencial do modelo - especifica a evolução do sistema entre dois instantes subseqüentes quaisquer. Para prever, é necessário relacionar os estados corrente e futuro através de uma relação probabilística $f(\theta_{i+1} | \theta_i, D_i)$ e para previsão e actualização da informação é preciso relacionar as observações futuras, digamos y_{i+n} , com o estado através da distribuição de probabilidade de y_{i+n} condicionada ao estado θ_{i+n} , $f(y_{i+n} | \theta_{i+n})$, ($n = 1, 2, \dots$). Esta definição sequencial permite que o método seja aplicável a um número pequeno de observações, além de facilitar a elaboração de um processo recursivo de estimação;

As previsões são obtidas em termos de distribuições de probabilidade.

Como já foi dito, a grande diferença entre as abordagens Clássica e Bayesiana reside na maneira pela qual os parâmetros do modelo são tratados. Enquanto, na abordagem clássica, estes são estimados via maximização da função de verosimilhança, na abordagem bayesiana, são designados subjectivamente ou sequencialmente estimados via inferência Bayesiana.

8.2 PRINCÍPIOS BÁSICOS DA INFERÊNCIA BAYESIANA

Em geral num problema de inferência indutiva (extensão do particular para o geral, nomeadamente da amostra para a população) as fontes de informação disponíveis são de dois tipos :

- (1) **A informação à priori**, anterior ou externa em relação à amostra ou à experiência, proveniente de toda a acumulação de conhecimentos sob o problema em estudo ou em situações afins, bem como da intuição ou ponto de vista subjectivo do analista ou de outros peritos sobre o assunto.
- (2) **A informação amostral** sob a forma de dados ou observações resultante de inquéritos, da repetição da experiência em condições constantes ou sensivelmente constantes ou, ainda dados históricos, como por exemplo no caso das sucessões cronológicas.

Como se sabe a inferência Clássica, de cuja escola se destacam nomes como R.A.Fisher, Jerry Neyman e E.S. Pearson, baseia-se exclusivamente na informação amostral. Os clássicos recusam a informação à priori alegando que esta não é processada formalmente na análise estatística, mesmo no caso em que essa informação é facilmente quantificável.

A inferência Bayesiana, de cuja escola se nomeiam F. Ramsey, De Finetti, H. Jeffreys, I.J.Good e D.V. Lindley, conjuga a informação à priori e a informação amostral através do Teorema de Bayes instrumento por excelência da análise Bayesiana, como veremos no decorrer desta segunda parte.

O conceito de probabilidade, ou melhor, a sua interpretação é outro ponto de controvérsia.

No que respeita ao aspecto formal da axiomática de Kolmogorov, não há grandes divergências. Os axiomas de Kolmogorov são um caso particular da axiomática de Rényi (1970) definidora da probabilidade condicional.

Porém quanto à ligação entre a função matemática, abstracta, e a interpretação do termo probabilidade quando este se refere a acontecimentos ou proposições com valor prático, é um ponto bastante polémico. Apenas de uma maneira muito sumária diremos que, do ponto de vista clássico, o conceito de probabilidade tem uma interpretação frequentista baseada no princípio de amostragem repetida, utilizando assim as frequências como medidas de incerteza. Do ponto de vista bayesiano a probabilidade é um conceito subjectivo que traduz o grau de crença pessoal na realização dos fenómenos aleatórios, sendo definida em termos de graus de credibilidade.

Do ponto de vista clássico a avaliação dos procedimentos estatísticos é feita em termos da frequência com que fornecem bons resultados, orientando-se pela precisão inicial ou pré-experimental. Na perspectiva da inferência Bayesiana é sustentado que a maioria das situações são não repetitivas (senão hipoteticamente) e portanto o que conta é a precisão obtida com o particular conjunto de dados disponível, preferindo portanto a precisão final ou pós-experimental. A propósito cite-se Berger (1980): *"When dealing with a one time situation, however, it is not clear what the relevance of initial precision is"*.

Partindo de axiomas que caracterizam a essência do comportamento racional, consistente e coerente de um indivíduo, em síntese, o denominado **princípio da coerência Bayesiana**, demonstra-se que os graus de crença podem ser expressos formalmente. Os problemas de inferência são estudados no quadro de um modelo probabilístico, onde existem parâmetros, θ , desconhecidos, que no modelo clássico constituem um escalar ou vector desconhecido, mas fixo. Para os bayesianos, tudo o que é desconhecido é incerto, passando assim o parâmetro a ser considerado como aleatório ou estocástico. A incerteza associada aos verdadeiros valores dos parâmetros do modelo passa pois a ser quantificada através de uma distribuição de probabilidade, (geralmente subjectiva) designada "**distribuição à priori**." Assim se θ é um parâmetro discreto, designando por $f(\theta)$ a função de probabilidade à priori, tem-se que $f(\theta)$

exprime o grau de credibilidade que o analista atribui ao valor particular do θ considerado; se θ é um parâmetro contínuo, designando por $f(\theta)$ a função de densidade de probabilidade (f. d. p.) à priori, tem-se que $f(\theta) d\theta$ exprime o grau de credibilidade que atribui ao intervalo $(\theta, \theta+d\theta)$.

Formalizando as considerações anteriores, considere-se uma variável ou vector aleatório observável representativo de uma ou N características de uma população ou universo. Seja x o valor observado de X .

As possíveis funções de probabilidade (f.p., caso discreto) ou funções de densidade de probabilidade (f.d.p., caso contínuo) de X designam-se por $f(x | \theta)$, onde θ , escalar ou vector é um parâmetro pertencente ao espaço do parâmetro Θ . O modelo probabilístico é assim constituído pela família $\mathcal{F} = \{ f(x | \theta) : \theta \in \Theta \}$, das distribuições probabilísticas representadas pela f.p. ou f.d.p. De futuro não se estará constantemente a distinguir entre o caso discreto e caso contínuo, exprimindo-se genericamente os raciocínios em termos da função de densidade $f(x | \theta)$.

Encarando o vector dos parâmetros, θ , como um vector aleatório (não se fará distinção de notação para o caso de θ fixo ou aleatório), com distribuição à priori $f(\theta)$, o teorema de Bayes para densidades permite escrever:

$$f(\theta|x) = \frac{f(\theta, x)}{f(x)} = \frac{f(\theta)f(x|\theta)}{\int_{\Theta} f(x|\theta) f(\theta) d\theta} \quad (8.1)$$

onde :

$f(\theta|x)$ representa a distribuição à posteriori de θ ;

$f(\theta, x)$ representa a distribuição conjunta das v.a. X e θ ;

$f(x)$ representa a distribuição marginal de X ;

$f(x|\theta)$ representa a função de verosimilhança (interpretada na abordagem bayesiana como a densidade de X condicional a θ);

\int_{Θ} representa o integral ou somatório estendido ao espaço paramétrico consoante este seja considerado contínuo ou discreto.

Note-se que ao designar-se por $f(\theta)$ a distribuição à priori de θ e por $f(x)$ a distribuição marginal de X , de modo algum se subentenda que θ e X têm a mesma distribuição.

Como a distribuição marginal de X não depende de θ , é usual escrever (8.1) sob a forma :

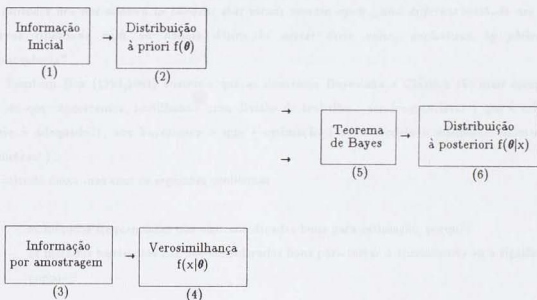
$$f(\theta|x) \propto f(\theta) f(x|\theta) \quad (8.2)$$

ou

distribuição à posteriori \propto dist. à priori \times verosimilhança da amostra ,

sendo a constante de proporcionalidade ou factor de normalização $K = \frac{1}{f(x)}$.

O **teorema de Bayes** permite assim modificar , ou melhor , rever a atitude inicial em relação a θ , traduzida por $f(\theta)$, tendo em conta a informação contida na amostra. Murteira (1988) esquematiza esse processo de revisão da distribuição à priori para obter a distribuição à posteriori da seguinte maneira :



Dada a forma de cálculo , os princípios nomeadamente da **suficiência** e da **verosimilhança** (formas fraca e forte) , são automaticamente satisfeitos.

Os clássicos opõem-se ao **princípio da verosimilhança** na medida em que defendem que a avaliação da informação fornecida pela amostra passa necessariamente pela análise de todas as amostras que poderiam ter ocorrido, isto é, todos os pontos do espaço amostra.

A distribuição marginal de X , $f(x)$, que foi utilizada no teorema de Bayes para a construção da distribuição à posteriori, tem interesse só por si , pois fornece a distribuição da amostra antes de ser observada , chamando-se por isso **distribuição preditiva**.

A introdução de informação à priori na análise estatística exige assim uma visão ampliada do conceito usual de probabilidade e da natureza do parâmetro desconhecido em causa.

Parece pois não ser de contestar que a abordagem bayesiana, ao fornecer um meio de introduzir informação adicional na análise, permite, na base de uma quantificação correcta dessa informação :

- Apurar muito possivelmente o procedimento inferencial sobre θ através do cálculo da sua distribuição condicionada pelos dados observados.
- Clarificar a sua interpretação, particularmente através da medição da precisão final, isto é, da precisão inferencial relativa à situação específica analisada.

Porém citando Barnett((1982) pg.188) "*... All this process of arguments is constructive provided that attitudes are not allowed to harden: that minds remain open , and different methods are used in different situations with an honest desire to assess their value, uncluttered by philosophical preconceptions*".

Também Box (1983,1984) sustenta que as doutrinas Bayesiana e Clássica são mais complementares do que concorrentes, perfilhando uma divisão de trabalho : aos frequentistas o que é crítica (o modelo é adequado?), aos bayesianos o que é estimação (se o modelo é adequado estimem-se os parâmetros!) .

Esta atitude deixa -nos com os seguintes problemas :

- os métodos frequentistas não são considerados bons para estimação, porquê?
- os métodos bayesianos não são considerados bons para testar o ajustamento ou a significância, porquê ?

A resposta não é pacífica . Não sendo este o objectivo do nosso trabalho , quedar-nos-emos por aqui.

8.3 DISTRIBUIÇÕES À PRIORI

8.3.1 Distribuições à priori não informativas .

Como se viu para obter a distribuição à posteriori (base de toda a inferência Bayesiana) é preciso especificar a distribuição à priori do parâmetro θ .

As distribuições à priori, em geral, classificam-se em dois tipos “ não informativas ” e “ informativas ”. As distribuições não informativas são as que traduzem a inexistência de informação inicial, isto é, uma situação de ignorância sobre o parâmetro em causa. Por vezes existe algum conhecimento, embora vago ou difuso, falando-se de *distribuições à priori vagas ou difusas* que por vezes se confundem com as distribuições não informativas (talvez porque na prática os resultados a que se chega sejam em geral muito semelhantes).

A especificação de distribuições à priori não informativas constitui um dos problemas mais controverso da análise bayesiana. Se o espaço dos parâmetros é finito, por exemplo $\Theta = \{ 1, 2, \dots, m \}$, uma distribuição à priori uniforme (discreta) $f(\theta) = \frac{1}{m}$; $\theta = 1, 2, \dots, m$, pode ser adequada para traduzir a ignorância à priori , pois de contrário estar-se-ia a considerar uns valores mais creíveis do que outros. A escolha da distribuição uniforme discreta baseia-se no princípio da razão insuficiente de Laplace (1812) - *Théorie Analytique des Probabilités*.

Quando Θ é um intervalo, $\Theta = \{ \theta : \theta_0 \leq \theta \leq \theta_1 \}$, será natural escolher para $f(\theta)$ a distribuição uniforme contínua no intervalo $[\theta_0, \theta_1]$.

Quando Θ é um intervalo não limitado, Jeffreys (1961) segue essencialmente o mesmo princípio , embora com ligeiras modificações. Alargando também o *princípio da razão insuficiente* para o caso multivariado, as sugestões de Jeffreys são :

Quando $\Theta = (-\infty, +\infty)$ e θ parâmetro de localização, deve tomar-se para distribuição à priori ,

$$f(\theta) d\theta \propto d\theta \quad (8.3)$$

Quando $\Theta = (0, +\infty)$ e θ parâmetro de escala , deve tomar-se ,

$$f(\theta) d\theta \propto \frac{1}{\theta} d\theta \quad (8.4)$$

Tanto (8.3) como (8.4) são distribuições impróprias pois que ,

$$\int_{-\infty}^{+\infty} d\theta = \infty \quad \text{e} \quad \int_0^{+\infty} \frac{1}{\theta} d\theta = \infty$$

Esta situação é criticada por outros autores, nomeadamente Box e Tiao, os quais introduzem o conceito de distribuição à priori localmente uniforme na região em que a função de verosimilhança tem um valor apreciável, mas recusam a sua extensão a todo o domínio do parâmetro.

Os seguidores de Jeffreys não se incomodam com o facto da distribuição à priori ser imprópria desde que a distribuição à posteriori resultante seja própria.

Jeffreys justifica a escolha daquelas distribuições à priori baseando-se nas propriedades de invariância para determinado tipo de transformações, o que lhe permite assegurar a consistência das conclusões à posteriori face às adequadas parametrizações alternativas. Por exemplo, se θ é um parâmetro de localização, a distribuição à priori deve ser invariante para transformações da forma $\lambda = \alpha + \beta\theta$, o que na realidade acontece pois que $d\lambda \propto d\theta$.

Generalizando as propriedades de invariância que justificam a escolha de (8.3) e (8.4), Jeffreys (1961) introduz uma especificação alternativa para as distribuições à priori,

$$f(\theta) \propto \{I(\theta)\}^{\frac{1}{2}}, \quad (8.5)$$

onde $I(\theta)$ é a quantidade de informação de Fisher,

$$I(\theta) = E \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right] = \int \left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 f(x|\theta) dx \quad \text{ou}$$

$$I(\theta) = -E \left[\left(\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right) \right] = - \int \left(\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right) f(x|\theta) dx \quad (8.5)$$

Se a parametrização do modelo for feita em termos de $\lambda = \tau(\theta)$, com τ função biunívoca e derivável, e se toma como distribuição à priori,

$$f(\lambda) \propto \{I(\lambda)\}^{\frac{1}{2}}, \quad (8.6)$$

então,

$$\{I(\lambda)\}^{\frac{1}{2}} d\lambda = \{I(\theta)\}^{\frac{1}{2}} d\theta \quad (8.7)$$

Significando assim que as distribuições à posteriori são análogas e consequentemente as inferências delas decorrentes serão equivalentes.

Por exemplo se $X \sim N(\mu, \sigma^2)$ e se supõe σ^2 conhecido, tem-se,

$$\log f(x|\mu) = \log(1/\sigma\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma^2} \quad e$$

$$\frac{\partial \log f(x|\mu)}{\partial \mu} = \frac{(x-\mu)}{\sigma^2}$$

donde,

$$I(\mu) = E \left[\left(\frac{\partial \log f(x|\mu)}{\partial \mu} \right)^2 \right] = \frac{1}{\sigma^2} = \text{constante}$$

ou seja $f(\mu) \propto \{I(\mu)\}^{\frac{1}{2}} = \text{constante}$, conduz-nos a (8.3).

Supondo, agora μ conhecido, e σ desconhecido, vem,

$$\frac{\partial \log f(x|\sigma)}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3} \quad e$$

$$\frac{\partial^2 \log f(x|\sigma)}{\partial \sigma^2} = \frac{1}{\sigma^2} - 3 \frac{(x-\mu)^2}{\sigma^4},$$

donde,

$$I(\sigma) = -E \left[\frac{1}{\sigma^2} - 3 \frac{(X-\mu)^2}{\sigma^4} \right] = \frac{2}{\sigma^2},$$

pelo que aplicando (8.6) é-se conduzido a (8.4), isto é,

$$f(\sigma) d\sigma \propto \frac{1}{\sigma} d\sigma.$$

Uma generalização de (8.6) para o caso multiparamétrico é feita usando em vez da quantidade de

informação, a matriz de informação de Fisher $\mathbf{I}(\theta)$, e definindo a distribuição à priori proporcional à raiz quadrada do respectivo determinante. Assim se θ for um parâmetro multivariado tem-se,

$$f(\theta) \propto |\mathbf{I}(\theta)|^{\frac{1}{2}},$$

onde,

$$\mathbf{I}(\theta) = -E \left[\left(\frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \partial \theta_j} \right) \right] \quad (8.8)$$

Por exemplo, para o caso da distribuição $N(\mu, \sigma^2)$ com ambos os parâmetros desconhecidos como,

$$\mathbf{I}(\mu, \sigma) = \begin{bmatrix} -E \left[\frac{\partial^2 \log f}{\partial \mu^2} \right] & -E \left[\frac{\partial^2 \log f}{\partial \mu \partial \sigma} \right] \\ -E \left[\frac{\partial^2 \log f}{\partial \sigma \partial \mu} \right] & -E \left[\frac{\partial^2 \log f}{\partial \sigma^2} \right] \end{bmatrix}$$

$$|\mathbf{I}(\mu, \sigma)|^{\frac{1}{2}} = \begin{vmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{vmatrix}^{\frac{1}{2}} = (2\sigma^{-4})^{\frac{1}{2}}$$

$$\text{assim, } f(\mu, \sigma) \propto \sigma^{-2} \quad (8.9)$$

A questão da averiguação do grau em que diferentes distribuições à priori alternativas, incorporam informação acerca de θ , pode ser respondida em termos da teoria da informação. A resposta não será obviamente única, atendendo às várias medidas de informação e definições de distribuição à priori de informação mínima, que podem ser utilizadas.

Bernardo (1979), seguindo Lindley (1956) usa a medida de informação logarítmica assim definida:

$$I^{\theta} \{ \epsilon, f(\theta) \} = \int f(x) \int f(\theta|x) \log \frac{f(\theta|x)}{f(\theta)} d\theta dx,$$

onde

$\epsilon = \{ X, \Theta, f(x|\theta) \}$ representa a experiência que produz uma observação de X , com distribuição $f(x|\theta)$ para algum $\theta \in \Theta$ e as restantes quantidades tendo o significado já atrás definido em (8.1).

Seja $I^\theta \{ \epsilon(N), f(\theta) \}$ a quantidade de informação sobre θ contida numa amostra de dimensão N . Se a dimensão da amostra fosse infinita obter-se-ia a quantidade de informação máxima sobre θ , $I^\theta \{ \epsilon(\infty), f(\theta) \}$. Assim,

$$G = (I^\theta \{ \epsilon(\infty), f(\theta) \} - I^\theta \{ \epsilon, f(\theta) \})$$

mede o ganho de informação acerca de θ quando a distribuição à priori de θ é $f(\theta)$. Uma distribuição à priori com "informação mínima" será a que maximiza G .

No caso contínuo, em geral $I^\theta \{ \epsilon(\infty), f(\theta) \} = +\infty$, para todos os $f(\theta)$. O que em parte é natural, pois que uma quantidade de informação infinita será necessária para conhecer exactamente um número real. Assim distribuição à priori de "informação mínima" poderá ser definida como a distribuição limite da sequência de distribuições que se vão obtendo à medida que N tende para infinito. Isto é, $f(\theta) = \lim f_N(\theta)$, onde $f_N(\theta)$ é a distribuição que maximiza $G_N = (I^\theta \{ \epsilon(N), f(\theta) \} - I^\theta \{ \epsilon, f(\theta) \})$. A correspondente distribuição à posteriori é obtida pelo teorema de Bayes. As distribuições à posteriori baseadas em distribuições à priori de informação mínima, também designadas distribuições à priori vagas ou difusas, costumam designar-se por "distribuições à posteriori de referência", pois elas podem constituir um ponto de partida admissível para a inferência.

Considerando o caso particular de populações Normais, $X \sim N(\mu, \sigma^2)$, com ambos os parâmetros desconhecidos, Bernardo(1979), deduz a distribuição à priori de informação mínima conjuntamente para μ e σ , obtendo:

$$f(\mu, \sigma) \propto \sigma^{-1} \quad (8.10)$$

Jeffreys, como vimos, a partir de (8.6) chega a $f(\mu, \sigma) \propto \sigma^{-2}$. Porém Jeffreys rejeita este resultado a favor de (8.10), pois sustenta que a independência à priori entre μ e σ deve ser de considerar, pelo que partindo das distribuições à priori $f(\sigma|\mu) \propto \sigma^{-1}$ e $f(\mu) \propto d\mu$, obtém (8.10).

As distribuições à posteriori de referência para μ e σ são respectivamente a distribuição t de Student com $(N-1)$ graus de liberdade e a distribuição do Qui-quadrado também com $(N-1)$ graus de liberdade, ou seja:

$$f(\mu | x_1, x_2, \dots, x_N) \propto [1 + \{(x - \mu) / s\}^2]^{-\frac{1}{2}N},$$

$$\text{onde } s^2 = \sum (x_i - \bar{x})^2 / N.$$

$$f(\sigma | x_1, x_2, \dots, x_N) \propto \sigma^{-N} \exp \{ -N s^2 / 2\sigma^2 \},$$

$$\text{isto é } N s^2 / \sigma^2 \sim \chi_{N-1}^2.$$

A designação de uma distribuição à priori não informativa pode ser embaraçosa dadas as várias técnicas propostas para a sua obtenção, não serem em geral concordantes.

8.3.2 Famílias Conjugadas de Distribuições .

Quando existe informação à priori substancial (relativamente à informação contida nos dados) é crucial assegurar que a distribuição à priori a utilizar constitua uma expressão precisa e completa dessa informação disponível, uma vez que as conclusões inferenciais vão depender da forma assumida pela distribuição à priori. Será também conveniente que haja meios de comparação entre as contribuições relativas das informações à priori e amostral, que o processo de passagem da distribuição à priori para a posteriori possa ser facilmente interpretado e que se faça de uma maneira "suave" sem que envolva cálculos de elevada dificuldade e morosos. Com o fim de conseguir estes objectivos Raiffa e Schlaifer (1961, 1972, 1977) introduziram o conceito de "famílias conjugadas de distribuições à priori".

Seja $\mathfrak{F} = \{f(x | \theta) : \theta \in \Theta\}$ a família considerada como modelo da situação prática em estudo, a classe $\mathfrak{G} = \{f(\theta) : \theta \in \Theta\}$ de distribuições à priori para θ é uma família conjugada em relação a \mathfrak{F} se e só se a distribuição à posteriori de θ , $f(\theta | x) \propto f(\theta) f(x | \theta)$, ainda pertence a \mathfrak{G} , quaisquer que sejam $f(x | \theta) \in \mathfrak{F}$ e $f(\theta) \in \mathfrak{G}$. Pode pois dizer-se que \mathfrak{G} é fechada em relação à operação de amostragem a partir de \mathfrak{F} .

Se existir estatística suficiente de dimensão fixa para \mathfrak{F} , caso da família exponencial, por exemplo, então existe família conjugada para \mathfrak{F} . Uma demonstração pode ver-se em Murteira (1988, pag. 91).

Para uma dada classe de densidades \mathfrak{F} , a família conjugada pode ser frequentemente determinada pela simples inspecção da função de verosimilhança $f(x | \theta)$, e escolher para a família conjugada, a classe de distribuições com a mesma forma funcional da função de verosimilhança. As distribuições à priori resultantes são designadas por *distribuições conjugadas naturais à priori*.

Como casos particulares de famílias conjugadas apresentam-se os ligados à família \mathfrak{F} de distribuições normais, que serão os que iremos utilizar nesta segunda parte do nosso trabalho.

(1) Distribuição Normal de média θ (desconhecida) e variância σ^2 (conhecida):

Admitindo à priori que $\theta \sim N(\mu, \tau^2)$, com μ e τ^2 também conhecidos, e que $f(x|\theta) \sim N(\theta, k\sigma^2)$, a função de densidade conjunta do par (X, θ) é dada por

$$f(x, \theta) = f(\theta) f(x|\theta) = (2\pi\sqrt{k}\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left[\left(\frac{\theta - \mu}{\tau} \right)^2 + \frac{(x - \theta)^2}{k\sigma^2} \right] \right\}$$

Definindo, $\rho = \frac{1}{\tau^2} + \frac{1}{k\sigma^2}$ e completando o quadrado do expoente, vem,

$$f(x, \theta) = (2\pi\sqrt{k}\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x}{k\sigma^2} \right) \right]^2 \right\} \exp \left\{ -\frac{(\mu - x)^2}{2(\tau^2 + k\sigma^2)} \right\}$$

A distribuição preditiva de X , obtém-se marginalizando θ ,

$$f(x) = \int_{-\infty}^{+\infty} f(x, \theta) d\theta = (2\pi\rho)^{-\frac{1}{2}} (\sqrt{k}\sigma\tau)^{-1} \exp \left\{ -\frac{(\mu - x)^2}{2(\tau^2 + k\sigma^2)} \right\} \quad (8.11)$$

A distribuição à posteriori de θ , aplicando o teorema de Bayes, vem

$$f(\theta|x) = \frac{f(x, \theta)}{f(x)} = \left(\frac{\rho}{2\pi} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x}{k\sigma^2} \right) \right]^2 \right\}$$

ou seja $f(\theta|x)$ é ainda uma distribuição Normal com média,

$$\begin{aligned} E(\theta|x) &= \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x}{k\sigma^2} \right) = \left\{ \left(\frac{1}{\tau^2} \right) \mu + \left(\frac{1}{k\sigma^2} \right) x \right\} / \rho = \frac{(1/\tau^2)\mu + (1/k\sigma^2)x}{(1/\tau^2) + (1/k\sigma^2)} = \\ &= \frac{k\sigma^2}{(\tau^2 + k\sigma^2)} \mu + \frac{\tau^2}{(\tau^2 + k\sigma^2)} x \end{aligned}$$

Somando e subtraindo $\frac{k\sigma^2}{(\tau^2 + k\sigma^2)} x$ vem,

$$\begin{aligned}
 E(\theta | x) &= \frac{k\sigma^2}{(\tau^2 + k\sigma^2)} \mu + \frac{\tau^2}{(\tau^2 + k\sigma^2)} x + \frac{k\sigma^2}{(\tau^2 + k\sigma^2)} x - \frac{k\sigma^2}{(\tau^2 + k\sigma^2)} x = \\
 &= x - \frac{k\sigma^2}{(\tau^2 + k\sigma^2)} (x - \mu)
 \end{aligned} \quad (8.12)$$

e variância,

$$\text{var}[\theta | x] = \rho^{-1} = \left(\frac{1}{\tau^2} + \frac{1}{k\sigma^2} \right)^{-1} \quad (8.13)$$

No caso de observar uma amostra aleatória de dimensão N , $x = (x_1, \dots, x_N)$, como se sabe a estatística suficiente é $\bar{X} \sim N(\theta, \sigma^2 / N)$ e portanto,

$f(\theta | x_1, \dots, x_N) \equiv f(\theta | \bar{x})$ ainda é Normal com média,

$$E[\theta | x_1, \dots, x_N] \equiv E[\theta | \bar{X}] = \left\{ \left(\frac{1}{\tau^2} \right) \mu + \left(\frac{1}{\sigma^2} \right) \bar{x} \right\} / \rho_N \quad (8.14)$$

e variância,

$$\text{var}[\theta | x_1, \dots, x_N] \equiv \text{var}[\theta | \bar{X}] = \rho_N^{-1} = \left(\frac{1}{\tau^2} + \frac{N}{\sigma^2} \right)^{-1} \quad (8.15)$$

Expressões que se obtêm respectivamente de (8.13) e (8.14) para o caso particular de $k = \frac{1}{N}$ e x substituído por \bar{x} .

Assim, a conjugada da Normal (variância conhecida) ainda é uma Normal.

(2) Distribuição Normal com média θ (conhecida) e variância σ^2 (desconhecida).

Considerando uma amostra $x = (x_1, \dots, x_N)$ de uma população $N(\theta, \sigma^2)$ e a parametrização em termos da precisão $\phi = \frac{1}{\sigma^2}$, a verosimilhança da amostra é como se sabe proporcional a,

$$\phi^{\frac{1}{2}N} \exp \left\{ -\frac{1}{2} \phi \sum (x_i - \theta)^2 \right\} = \phi^{\frac{1}{2}N} \exp \left\{ -\frac{1}{2} \phi N s^* \right\},$$

onde $s^* = \frac{1}{N} \sum (x_i - \theta)^2$ e' estatística suficiente.

Admitindo que a distribuição à priori de ϕ é uma Gama com parâmetros $\frac{\alpha_0}{2}$ e $\frac{\beta_0}{2}$, com α_0 e β_0 positivos (consideram-se parâmetros $\frac{\alpha_0}{2}$ e $\frac{\beta_0}{2}$ para comodidade de cálculos e também porque assim $\beta_0\phi$ tem distribuição qui-quadrado com α_0 graus de liberdade).

$$f(\phi) \equiv f(\phi, \alpha_0, \beta_0) \propto \phi^{\left(\frac{\alpha_0}{2}-1\right)} \exp \left\{ -\frac{1}{2} \beta_0 \phi \right\}, \quad (8.16)$$

então a distribuição à posteriori, sendo proporcional ao produto da densidade à priori pela verossimilhança vem,

$$f(\phi | x) \propto \phi^{\frac{1}{2}(\alpha_0 + N) - 1} \exp \left\{ -\frac{1}{2} \phi [\beta_0 + N s^*] \right\} \quad (8.17)$$

isto é a distribuição à posteriori ainda é Gama com parâmetros ,

$$\alpha_1 = \alpha_0 + N \quad (8.18)$$

$$\beta_1 = \beta_0 + N s^* = \beta_0 + \sum (x_i - \theta)^2 \quad (8.19)$$

Assim a conjugada da Normal (média conhecida) é uma Gama.

Se a parametrização for feita em termos da variância a conjugada é a Gama inversa (veja-se Raiffa e Schlaifer (1977)).

(3) Distribuição Normal de média θ e variância σ^2 (ambas desconhecidas).

Continuando a considerar a parametrização em termos da precisão $\phi = \frac{1}{\sigma^2}$ e admitindo que $f(\phi) \sim G\left(\frac{\alpha_0}{2}, \frac{\beta_0}{2}\right)$ e $f(\theta | \phi) \sim N(m_0, (N_0 \phi)^{-1})$ a distribuição conjunta de (θ, ϕ) , $f(\theta, \phi) = f(\theta | \phi) f(\phi)$, é o produto de uma distribuição Normal por uma distribuição Gama sendo, por isso, designada por distribuição Normal - Gama. Assim,

$$f(\theta | \phi) \propto \phi^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} N_0 \phi (\theta - m_0)^2 \right\}$$

$$f(\phi) \propto \phi^{\left(\frac{\alpha_0}{2}-1\right)} \exp \left\{ -\frac{1}{2} \beta_0 \phi \right\},$$

pelo que a distribuição conjunta à priori de (θ, ϕ) vem,

$$f(\theta, \phi) \propto \phi^{\frac{1}{2}(\alpha_0 + 1) - 1} \exp \left\{ -\frac{1}{2} N_0 \phi (\theta - m_0)^2 \right\} \exp \left\{ -\frac{1}{2} \beta_0 \phi \right\}, \quad (8.20)$$

isto é, $f(\theta, \phi) \sim \text{NG} (m_0, N_0, \alpha_0, \beta_0)$.

Considerando uma amostra $x = (x_1, \dots, x_N)$ ($f(x_i|\theta, \phi) \sim N(\theta, \sigma^2 = \phi^{-1})$) a verossimilhança da amostra é proporcional a ,

$$\phi^{\frac{1}{2}N} \exp \left\{ \sum \left[-\frac{1}{2} \phi (x_i - \theta)^2 \right] \right\} = \phi^{\frac{1}{2}N} \exp \left\{ \left[-\frac{1}{2} N \phi (\bar{x} - \theta)^2 \right] \exp \left\{ -\frac{1}{2} \phi (N-1) s'^2 \right\} \right\}$$

onde $s'^2 = \frac{\sum (x_i - \bar{x})^2}{N-1}$, pois que ,

$$\sum (x_i - \theta)^2 = N (\bar{x} - \theta)^2 + (N-1) s'^2.$$

A densidade à posteriori é proporcional ao produto da densidade à priori pela verossimilhança ; associando convenientemente os factores vem :

$$f(\theta, \phi | x) \propto \phi^{\frac{1}{2}(N + \alpha_0 + 1) - 1} \exp \left\{ -\frac{1}{2} \phi [N (\bar{x} - \theta)^2 + N_0 (\theta - m_0)^2] \right\} \cdot \exp \left\{ -\frac{1}{2} \phi [\beta_0 + (N-1) s'^2] \right\} \quad (8.21)$$

Como $N (\bar{x} - \theta)^2 + N_0 (\theta - m_0)^2 = (N + N_0) \theta^2 - 2\theta (N \bar{x} + N_0 m_0) + N \bar{x}^2 + N_0 m_0^2 =$

$$= (N + N_0) \left\{ \theta^2 - 2\theta (N \bar{x} + N_0 m_0)/(N + N_0) + (N \bar{x}^2 + N_0 m_0^2)/(N + N_0) \right\} \quad (8.22)$$

Definindo,

$$m_1 = (N \bar{x} + N_0 m_0)/(N + N_0) \quad (8.23)$$

$$N_1 = N + N_0 \quad (8.24)$$

$$\alpha_1 = N + \alpha_0 \quad (8.25)$$

e completando o quadrado em θ na expressão (8.22) (basta somar e subtrair m_1^2) vem,

$$f(\theta, \phi | x) \propto \phi^{\frac{1}{2}(\alpha_1 + 1) - 1} \exp \left\{ -\frac{1}{2} \phi N_1 (\theta - m_1)^2 \right\} \exp \left\{ -\frac{1}{2} \phi [(N-1) s'^2 + \beta_0] \right\} -$$

$$- \frac{(N \bar{x} + N_0 m_0)^2}{N_1} + (N \bar{x}^2 + N_0 m_0^2)] \} ,$$

como

$$\begin{aligned} [(N-1) s'^2 + \beta_0] - \frac{(N \bar{x} + N_0 m_0)^2}{N_1} + (N \bar{x}^2 + N_0 m_0^2) &= \\ = [(N-1) s'^2 + N \bar{x}^2 + (\beta_0 + N_0 m_0^2) - N_1 m_1^2] , \end{aligned}$$

definindo,

$$\beta_1 = [(N-1) s'^2 + N \bar{x}^2] + (\beta_0 + N_0 m_0^2) - N_1 m_1^2 \quad (8.26)$$

tem-se que ,

$$f(\theta, \phi | x) \propto \phi^{\frac{1}{2}(\alpha_1 + 1) - 1} \exp \{ -\frac{1}{2} \phi N_1 (\theta - m_1)^2 \} \exp \{ (-\frac{1}{2} \phi \beta_1) \} . \quad (8.27)$$

Pelo que a distribuição à posteriori ainda é uma Normal - Gama ,

$$f(\theta, \phi | x) \sim NG (m_1 , N_1 , \alpha_1 , \beta_1)$$

com parâmetros m_1 , N_1 , α_1 e β_1 definidos respectivamente por (8. 23) , (8.24) , (8.25) e (8.26) .

8.4 ESTIMAÇÃO BAYESIANA.

A distribuição à posteriori, como se viu, incorpora toda a informação disponível sobre o parâmetro (informação inicial mais informação amostral) pelo que todos os procedimentos de inferência Bayesiana são baseados na distribuição à posteriori, $f(\theta | x)$, que em paralelo com a inferência clássica se pode dizer tem um duplo papel substituindo a f. de verosimilhança e as distribuições por amostragem. A distribuição à posteriori, além de fornecer os procedimentos (processos de estimação, testes de hipóteses, etc) permite ainda avaliar a qualidade desses procedimentos. Assim na inferência clássica, a estimativa pontual de θ é o valor $\hat{\theta}$, que maximiza a função de verosimilhança, na estimação bayesiana em paralelo teremos o valor de θ , $\hat{\theta}$, que torna máxima a distribuição à posteriori, i.e. obter-se-á, assim a moda, interpretada como o valor mais provável de θ , na situação específica em causa (dada a distribuição à priori e a amostra). A moda da distribuição à posteriori é aleatória antes de se dispor da amostra particular, e

portanto, é considerada o estimador Bayesiano de θ .

Se passarmos do campo estritamente inferencial, incorporando informação adicional sobre as consequências da selecção incorrecta do valor do parâmetro $\theta \in \Theta$, é possível definir outras alternativas. Um critério largamente utilizado no âmbito da teoria da decisão é o de escolher a estimativa que minimiza a perda esperada à posteriori. Seja $L(\theta, \theta^*)$ a função perda então,

$$\min_{\theta^*} E [L(\theta, \theta^*)] = \min_{\theta^*} \int_{\Theta} L(\theta, \theta^*) f(\theta | x) d\theta \quad (8.28)$$

De acordo com este critério é fácil provar que a estimativa óptima (no sentido definido) para funções perda quadráticas definidas positivamente é a média à posteriori, caso exista. Na primeira parte deste trabalho este resultado foi demonstrado tomando para função perda o E.Q.M. Naturalmente que o uso de outros tipos de função perda levará a outras estimativas óptimas, a moda ou a mediana à posteriori, por exemplo.

Como se sabe no caso da distribuição à posteriori ser Normal, como é unimodal e simétrica, os resultados obtidos para o parâmetro média pelos dois critérios mencionados são coincidentes, pois a média, a mediana e a moda coincidem.

8.5 REGIÕES DE CONFIANÇA E TESTES DE HIPÓTESES BAYESIANOS.

Para efeitos práticos informação à posteriori sobre θ , pode ser obtida através de intervalos ou regiões com um valor de probabilidade fixada. Isto é, $R_{\alpha}(x)$ é uma região de confiança Bayesiana ou um conjunto de credibilidade $(1-\alpha)$ se e só se,

$$P \{ \theta \in R_{\alpha}(x) \} = \int_{R_{\alpha}(x)} f(\theta | x) d\theta = 1 - \alpha \quad (8.29)$$

Tal como acontece com as regiões de confiança clássicas, a região $R_{\alpha}(x)$ para um dado α , não é única, sendo também de desejar que abranja o menor volume possível no espaço paramétrico. Assim $R_{\alpha}(x)$ é uma região com a mais elevada probabilidade, abreviadamente uma região HPD (" Highest Posteriori Density ") se

$$f(\theta_1 | x) \geq f(\theta_2 | x), \forall \theta_1 \in R_{\alpha}(x) \text{ e } \forall \theta_2 \notin R_{\alpha}(x),$$

HPD regiões constituem assim a base para a inferência Bayesiana (estimação e testes de hipóteses) .

O problema dos testes de hipóteses pode colocar-se muito sumariamente da seguinte maneira . Suponha-se uma partição $\{ \Theta_0, \Theta_1 \}$ do espaço paramétrico , isto é, $\Theta = \Theta_0 \cup \Theta_1$ e Θ_0 e Θ_1 disjuntos e sejam H_0 e H_1 as duas hipóteses sobre o parâmetro θ , mutuamente exclusivas e exaustivas :

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1 .$$

Considerando as “ chances ” à posteriori associadas a cada uma das hipóteses ,

$$\alpha_0 = p(H_0 | x) = \int_{\theta \in \Theta_0} f(\theta | x) d\theta$$

$$\alpha_1 = p(H_1 | x) = \int_{\theta \in \Theta_1} f(\theta | x) d\theta,$$

decide-se entre H_0 e H_1 de acordo com os valores de α_0 e α_1 , optando pela hipótese com maior probabilidade à posteriori ou equivalentemente considerando o cociente , α_0 / α_1 , designado por “ *razão de chances à posteriori* ” optar-se-á por H_0 se aquele for > 1 .

Uma medida de interesse para a análise Bayesiana é o designado “ *factor de Bayes* ” que é definido pelo cociente entre as razões de chances à posteriori e à priori , isto é ,

$$B = \frac{\text{“ razão de chances à posteriori ”}}{\text{“ razão de chances à priori ”}} = \frac{\alpha_0 / \alpha_1}{\pi_0 / \pi_1} ,$$

onde π_0 e π_1 designam respectivamente as probabilidades à priori das hipóteses H_0 e H_1 .

O interesse do factor de Bayes reside no facto de este poder, algumas vezes, ser interpretado como “ as chances de H_0 sobre H_1 ” fornecidas pelos dados. Esta é uma interpretação claramente válida quando as hipóteses são simples, i.e, $\Theta_0 = \{ \theta_0 \}$ e $\Theta_1 = \{ \theta_1 \}$,

$$\alpha_0 = p(H_0 | x) = \frac{\pi_0 f(x | \theta_0)}{\pi_0 f(x | \theta_0) + \pi_1 f(x | \theta_1)}$$

$$\alpha_1 = p(H_1 | x) = \frac{\pi_1 f(x | \theta_1)}{\pi_0 f(x | \theta_0) + \pi_1 f(x | \theta_1)},$$

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(x | \theta_0)}{\pi_1 f(x | \theta_1)} \quad \text{e} \quad B = \frac{\alpha_0 \pi_0}{\alpha_1 \pi_1} = \frac{f(x | \theta_0)}{f(x | \theta_1)}.$$

Assim podemos dizer que o factor de Bayes é precisamente a razão de verosimilhanças de H_0 para H_1 .

Consideremos agora duas distribuições conjuntas de informações, digamos f , que passamos agora a considerar por H_0 , por um lado, de modo a considerar as observações $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ em função da informação θ priori, H_0 , e da mesma em particular as variáveis aleatórias de variáveis de observação $H_1 = (X_1, Y_1, \dots, X_n, Y_n, H_1)$ de H_0 . Assim se designarmos por H_1 toda a variável de informação dependente de H_0 e θ , considerando que a probabilidade f é dada para a informação externa H_0 pode designar-se $f(H_1)$, tornando-se $H_1 = (X_1, Y_1, \dots, X_n, Y_n)$.

Seja g a distribuição, segundo a primeira distribuição, da variável bayesiana, e nos comportamentos deve ser representada em termos de uma distribuição de probabilidade, condicionada pela informação de H_0 , $g(H_1 | H_0)$, distribuição g condicionada por H_0 .

Concluamos, para que a distribuição da teoria de Bayes nos forneça a distribuição g posterior, $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$.

$$X_1 = Y_1, X_2 = Y_2, \dots, X_n = Y_n \quad (1.1)$$

$$X_1 = Y_1, X_2 = Y_2, \dots, X_n = Y_n \quad (1.2)$$

onde se continua a supor que os variáveis X_1, X_2, \dots, X_n são independentes e mutuamente independentes com as variáveis Y_1, Y_2, \dots, Y_n . Assim, para uma distribuição g posterior, $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$, a distribuição g condicionada por H_0 , $g(H_1 | H_0)$.

$$\begin{aligned} g(H_1 | H_0) &= \int g(H_1, Y_1, \dots, Y_n | H_0) dY_1 \dots dY_n \\ &= \int g(H_1, Y_1, \dots, Y_n | H_0) dY_1 \dots dY_n \end{aligned} \quad (1.3)$$

A expressão da distribuição g posterior, $g(H_1 | H_0)$, é dada por:

$$g(H_1 | H_0) = g(H_1 | H_0, Y_1, \dots, Y_n) = \frac{g(H_1 | H_0) g(Y_1 | H_0, Y_2, \dots, Y_n)}{g(Y_1 | H_0, Y_2, \dots, Y_n)}$$

9. MODELOS ESTRUTURAIS BAYESIANOS - - FORMULAÇÃO MATEMÁTICA

A formulação matemática dos modelos estruturais bayesianos é idêntica à apresentada na primeira parte deste trabalho. Retomemos pois, o MLD definido pelas equações (4.1) e (4.2). Em cada instante t o processo em estudo é caracterizado por um conjunto de parâmetros, θ_t , desconhecidos que se pretendem estimar face a determinado conjunto de informação disponível, que passaremos agora, a designar por D_t , pois que além de conter o conjunto das observações $Y_t = \{y_0, y_1, \dots, y_t\}$ contém todo o conjunto de informação à priori, D_0 , incluindo em particular os valores das matrizes definidoras do vector de caracterização $M_t = \{F_t, G_t, H_t, R_t, Q_t\}$ do MLD. Assim se designarmos por D_{t-1} todo o conjunto de informação disponível no instante $t-1$ e considerando que o modelo é fechado para a informação externa depois do instante $t=0$, teremos que $D_t = \{y_t, D_{t-1}\}$.

Sendo θ_t desconhecido, segundo a permissa fundamental da estatística bayesiana, o seu comportamento deve ser especificado em termos de uma distribuição de probabilidade, condicionada pela informação no instante t , i.e., $f(\theta_t | D_t)$, distribuição à posteriori no instante t .

Como vimos, para que o mecanismo do teorema de Bayes nos forneça a distribuição à posteriori, é necessário conhecer a distribuição à priori (antes da observação y_t), digamos $f(\theta_t | D_{t-1})$. Formalizando seja:

$$y_t = F_t \theta_t + \epsilon_t \quad (9.1)$$

$$\theta_t = G_{t-1} \theta_{t-1} + R_t \eta_t \quad (9.2)$$

onde se continua a admitir que os erros ϵ_t e η_t são independentes e mutuamente independentes com distribuições conhecidas e que o vector de estado inicial, θ_0 , também tem uma distribuição à priori conhecida, $f(\theta_0 | D_0)$. Assim a distribuição de $(\theta_t | \theta_{t-1})$ pode obter-se directamente da distribuição assumida para η_t e por conseguinte, ter-se-á:

$$\begin{aligned} f(\theta_t | D_{t-1}) &= \int f(\theta_t, \theta_{t-1} | D_{t-1}) d\theta_{t-1} = \\ &= \int f(\theta_t | \theta_{t-1}) f(\theta_{t-1} | D_{t-1}) d\theta_{t-1} \end{aligned} \quad (9.3)$$

As equações de actualização ou distribuições à posteriori obter-se-ão a partir de,

$$f(\theta_t | D_t) = f(\theta_t | y_t, D_{t-1}) = \frac{f(\theta_t | D_{t-1}) f(y_t | \theta_t, D_{t-1})}{f(y_t | D_{t-1})}$$

Como $f(\mathbf{y}_t | \boldsymbol{\theta}_t, D_{t-1}) = f(\mathbf{y}_t | \boldsymbol{\theta}_t)$ vem,

$$f(\boldsymbol{\theta}_t | D_t) = \frac{f(\boldsymbol{\theta}_t | D_{t-1}) f(\mathbf{y}_t | \boldsymbol{\theta}_t)}{f(\mathbf{y}_t | D_{t-1})} \quad (9.4)$$

onde a distribuição preditiva de \mathbf{y}_t se pode obter de ,

$$f(\mathbf{y}_t | D_{t-1}) = \int f(\mathbf{y}_t | \boldsymbol{\theta}_t) f(\boldsymbol{\theta}_t | D_{t-1}) d\boldsymbol{\theta}_t \quad (9.5)$$

deduzindo-se $f(\mathbf{y}_t | \boldsymbol{\theta}_t)$ directamente da equação das observações (9.1) a partir da distribuição das perturbações ϵ_t .

Como já se disse uma solução analítica para (9.5) pode não ser facilmente obtida tudo dependerá da forma das distribuições assumidas para ϵ_t , η_t e $\boldsymbol{\theta}_0$.

9.1 MODELO NORMAL LINEAR DINÂMICO MULTIVARIADO.

Considere-se o modelo definido por (9.1) e (9.2) com as hipóteses adicionais,

$$\epsilon_t \sim N_N(0, \mathbf{H}_t) \quad (9.6)$$

$$\eta_t \sim N_q(0, \mathbf{Q}_t) \quad (9.7)$$

$$\text{e } f(\boldsymbol{\theta}_0 | D_0) \sim N_m(\mathbf{a}_0, \mathbf{P}_0) \quad (9.8)$$

Nestas condições a distribuição à posteriori de $(\boldsymbol{\theta}_t | D_t)$ ainda é Normal com média que continuamos a designar por $\mathbf{a}_{t|t}$ e matriz de variâncias covariâncias $\mathbf{P}_{t|t}$ definidas como em (4.7) e (4.8) (equações definidoras do filtro de Kalman).

Na primeira parte deste trabalho este resultado foi obtido por indução e usando as propriedades da teoria da distribuição Normal multivariada. Vamos agora apresentar uma dedução baseada no teorema de Bayes. Antes porém, note-se que qualquer instante t pode ser considerado como instante zero, desde que se admita que $\mathbf{y}_1, \mathbf{y}_2, \dots$ representa a continuação da série observada previamente. Considerando-se

pois que a distribuição à priori concentra toda a informação do passado (actuando como se se tratasse de uma estatística suficiente), e interpretando-se θ_0 como o vector de estado final para os dados históricos.

Admita-se então que $(\theta_{t-1} | D_{t-1}) \sim N_m(\mathbf{a}_{t-1}, \mathbf{P}_{t-1})$, de (9.4) podemos escrever,

$$f(\theta_t | D_t) \propto f(\theta_t | D_{t-1}) f(y_t | \theta_t)$$

Da equação das observações (9.1) conjuntamente com a hipótese de normalidade de ϵ_t , obtem-se,

$$(y_t | \theta_t) \sim N_N(\mathbf{F}_t \theta_t, \mathbf{H}_t)$$

donde,

$$f(y_t | \theta_t) \propto \exp \left\{ -\frac{1}{2} [(y_t - \mathbf{F}_t \theta_t)^T \mathbf{H}_t^{-1} (y_t - \mathbf{F}_t \theta_t)] \right\} \quad (9.9)$$

Da equação de transição conjuntamente com as hipóteses sobre a distribuição inicial de $(\theta_{t-1} | D_{t-1})$ e de η_t , tem-se que $(\theta_t | D_{t-1})$ tem distribuição Normal de média,

$$\mathbf{a}_{t|t-1} = E[\theta_t | D_{t-1}] = E[(\mathbf{G}_{t-1} \theta_{t-1} + \mathbf{R}_t \eta_t) | D_{t-1}] = \mathbf{G}_{t-1} E[\theta_{t-1} | D_{t-1}]$$

$$\mathbf{a}_{t|t-1} = \mathbf{G}_{t-1} \mathbf{a}_{t-1}$$

e covariância,

$$\begin{aligned} \mathbf{P}_{t|t-1} &= E[\theta_t \theta_t^T | D_{t-1}] - E[\theta_t | D_{t-1}] E[\theta_t^T | D_{t-1}] = \\ &= E[(\mathbf{G}_{t-1} \theta_{t-1} + \mathbf{R}_t \eta_t) (\mathbf{G}_{t-1} \theta_{t-1} + \mathbf{R}_t \eta_t)^T | D_{t-1}] - \mathbf{G}_{t-1} \mathbf{a}_{t-1} \mathbf{a}_{t-1}^T \mathbf{G}_{t-1}^T = \\ &= \mathbf{G}_{t-1} (E[\theta_{t-1} \theta_{t-1}^T | D_{t-1}] - \mathbf{a}_{t-1} \mathbf{a}_{t-1}^T) \mathbf{G}_{t-1}^T + E[\mathbf{R}_t \eta_t \eta_t^T \mathbf{R}_t^T | D_{t-1}] = \\ &= \mathbf{G}_{t-1} \mathbf{P}_{t-1} \mathbf{G}_{t-1}^T + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t^T \end{aligned}$$

pois que η_s e θ_t são estatisticamente independentes, para $t < s$. Então,

$$f(\theta_t | D_{t-1}) \propto \exp \left\{ -\frac{1}{2} (\theta_t - \mathbf{a}_{t|t-1})^T \mathbf{P}_{t|t-1}^{-1} (\theta_t - \mathbf{a}_{t|t-1}) \right\} \quad (9.10)$$

Multiplicando termo a termo (9.10) e (9.9) obtém-se ,

$$f(\theta_t | D_t) \propto \exp \left\{ -\frac{1}{2} [(\theta_t - \mathbf{a}_{t|t-1})^T \mathbf{P}_{t|t-1}^{-1} (\theta_t - \mathbf{a}_{t|t-1}) + (\mathbf{y}_t - \mathbf{F}_t \theta_t)^T \mathbf{H}_t^{-1} (\mathbf{y}_t - \mathbf{F}_t \theta_t)] \right\}$$

Desenvolvendo o argumento da exponencial vem ,

$$\begin{aligned} & \theta_t^T \mathbf{P}_{t|t-1}^{-1} \theta_t - \theta_t^T \mathbf{P}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} - \mathbf{a}_{t|t-1}^T \mathbf{P}_{t|t-1}^{-1} \theta_t + \mathbf{a}_{t|t-1}^T \mathbf{P}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} + \\ & + \mathbf{y}_t^T \mathbf{H}_t^{-1} \mathbf{y}_t - \mathbf{y}_t^T \mathbf{H}_t^{-1} \mathbf{F}_t \theta_t - \theta_t^T \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{y}_t + \theta_t^T \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{F}_t \theta_t = \\ & = \theta_t^T \left[\mathbf{P}_{t|t-1}^{-1} + \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{F}_t \right] \theta_t - 2 \theta_t^T \left[\mathbf{P}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} + \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{y}_t \right] + c_1 \end{aligned} \quad (9.11)$$

A constante $c_1 = \mathbf{y}_t^T \mathbf{H}_t^{-1} \mathbf{y}_t + \mathbf{a}_{t|t-1}^T \mathbf{P}_{t|t-1}^{-1} \mathbf{a}_{t|t-1}$, não depende de θ_t .

Definindo ,

$$\mathbf{P}_{t|t}^{-1} = \mathbf{P}_{t|t-1}^{-1} + \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{F}_t \quad (9.12)$$

e

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{A}_t \nu_t \quad (9.13)$$

onde,

$$\mathbf{A}_t = \mathbf{P}_{t|t-1} \mathbf{F}_t^T \mathbf{Z}_t^{-1} \quad , \quad \mathbf{Z}_t = [\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{H}_t] \quad \text{e} \quad \nu_t = \mathbf{y}_t - \mathbf{F}_t \mathbf{a}_{t|t-1} ,$$

e após alguns cálculos, segue-se que:

$$\mathbf{P}_{t|t}^{-1} \mathbf{a}_{t|t} = \mathbf{P}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} + \mathbf{F}_t^T \mathbf{H}_t^{-1} \mathbf{y}_t$$

completando a forma quadrática (9.11) em θ_t , para o que basta somar e subtrair $\mathbf{a}_{t|t}^T \mathbf{P}_{t|t}^{-1} \mathbf{a}_{t|t}$, obtém-se para o argumento da exponencial

$$(\theta_t - \mathbf{a}_{t|t})^T \mathbf{P}_{t|t}^{-1} (\theta_t - \mathbf{a}_{t|t}) + c_2$$

onde $c_2 = c_1 - \mathbf{a}_{t|t}^T \mathbf{P}_{t|t}^{-1} \mathbf{a}_{t|t}$, também não depende de θ_t , pelo que

$$f(\theta_t | D_t) \propto \exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t})^T P_{t|t}^{-1} (\theta_t - a_{t|t}) \right\} \quad (9.14)$$

o que significa que a distribuição à posteriori de θ_t é Normal de média $a_{t|t}$ e matriz de variâncias covariâncias $P_{t|t}$.

Fazendo uso do seguinte lemma sobre matrizes (Lindley e Smith (1972)) : para quaisquer matrizes A , B e C de dimensões apropriadas e para as quais as matrizes B^{-1} e C^{-1} se admitem existir, tem-se

$$[B + A C A^T]^{-1} = B^{-1} - B^{-1} A [A^T B^{-1} A + C^{-1}]^{-1} A^T B^{-1},$$

como,

$$P_{t|t} = [P_{t|t-1}^T + F_t^T H_t^T F_t]^{-1}, \text{ aplicando o lemma vem :}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} F_t^T [F_t P_{t|t-1} F_t^T + H_t]^{-1} F_t P_{t|t-1}$$

que é precisamente a expressão (4.8) como se pretendia provar. Como a expressão (9.13) definidora de $a_{t|t}$ é exactamente a equação (4.7) obtivemos assim, numa óptica estritamente bayesiana, as equações de actualização do Filtro de Kalman.

Quanto à previsão a n passos tem-se que :

$$(\theta_{t+n} | D_t) \sim N_m(a_{t+n|t}, P_{t+n|t}) \quad e$$

$$(y_{t+n} | D_t) \sim N_N(\hat{y}_{t+n|t}, Z_{t+n|t})$$

onde $a_{t+n|t}$, $P_{t+n|t}$, $\hat{y}_{t+n|t}$ e $Z_{t+n|t}$ são obtidos respectivamente a partir de (4.27), (4.28), (4.31) e (4.32). Com efeito θ_{t+n} , exprime-se em função de θ_t a partir da expressão (4.26), isto é,

$$\theta_{t+n} = \left(\prod_{j=0}^{n-1} G_{t+j} \right) \theta_t + \sum_{j=1}^{n-1} \left(\prod_{i=j}^{n-1} G_{t+i} \right) R_{t+j} \eta_{t+j} + R_{t+n} \eta_{t+n} \quad (9.15)$$

Assim θ_{t+n} é uma combinação linear de variáveis normais independentes, pelo que também é normal,

$$(\theta_{t+n} | D_t) \sim N_m(a_{t+n|t}, P_{t+n|t})$$

As expressões para $a_{t+n|t}$ e $P_{t+n|t}$ foram deduzidas na primeira parte (cap.4 pag 35 e 36).

Quanto à distribuição de $\mathbf{y}_{t+n} | D_t$, tem-se que a equação das observações para $t=t+n$ é,

$$\mathbf{y}_{t+n} = \mathbf{F}_{t+n} \boldsymbol{\theta}_{t+n} + \boldsymbol{\epsilon}_{t+n}$$

pelo que, também devido à linearidade e independência, se tem ,

$$(\mathbf{y}_{t+n} | D_t) \sim N_N (\mathbf{F}_{t+n} \mathbf{a}_{t+n|t}, \mathbf{Z}_{t+n|t}) , \text{ onde } \mathbf{Z}_{t+n} = \mathbf{F}_{t+n} \mathbf{P}_{t+n|t} \mathbf{F}_{t+n}^T + \mathbf{H}_{t+n}$$

9.2 MODELO NORMAL LINEAR DINÂMICO UNIVARIADO.

O modelo univariado é, como é evidente, um caso particular do modelo multivariado, por isso tudo o que foi dito em (9.1) se transporta para o caso univariado, exactamente como já foi feito na primeira parte passando portanto \mathbf{y}_t a representar apenas uma variável , a matriz \mathbf{F}_t passará a ser um vector transposto (1xm) , \mathbf{f}_t , e a matriz \mathbf{H}_t (NxN) , das covariâncias do vector das perturbações associadas às observações passa a ser um escalar, h_t , assim como a matriz \mathbf{Z}_t (NxN) das covariâncias do vector das inovações , $\boldsymbol{\nu}_t$, passará a escalar , z_t . Assim teremos :

$$\text{Equação das observações : } y_t = f_t \boldsymbol{\theta}_t + \epsilon_t ; \epsilon_t \sim N (0 , h_t) \quad (9.16)$$

$$\text{Equação de transição : } \boldsymbol{\theta}_t = \mathbf{G}_{t-1} \boldsymbol{\theta}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t ; \boldsymbol{\eta}_t \sim N_q (0 , \mathbf{Q}_t) \quad (9.17)$$

$$\text{Distribuição à priori : } (\boldsymbol{\theta}_0 | D_0) \sim N_m (\mathbf{a}_0 , \mathbf{P}_0) \quad (9.18)$$

Mantendo-se as hipóteses de independência e independência mútua das perturbações ϵ_t e $\boldsymbol{\eta}_t$, bem como a independência de $(\boldsymbol{\theta}_0 | D_0)$ relativamente àquelas.

Tal como na primeira parte, até aqui , admitiu-se que o vector de caracterização, $\mathbf{M}_t = \{ \mathbf{f}_t , \mathbf{G}_t , \mathbf{R}_t , h_t , \mathbf{Q}_t \}$ era completamente conhecido para todo t . Esta situação é porém pouco plausível . Na prática e em geral, a variância , h_t , do processo gerador das observações y_t , não é conhecida, assim como a matriz das variâncias covariâncias, \mathbf{Q}_t . Trataremos em primeiro lugar o problema da estimação da variância das observações. Na abordagem clássica a sua estimação foi feita pelo método de máxima verosimilhança. Vejamos agora como na abordagem bayesiana este problema é resolvido, fazendo uso, de uma análise baseada em distribuições à priori conjugadas e depois como

análise de referência em distribuições vagas ou difusas.

9.2.1 INFERÊNCIA SOBRE A VARIÂNCIA DAS OBSERVAÇÕES, h_t .

Uso de famílias conjugadas de distribuições à priori

Admita-se que a variância, h_t é desconhecida, mas invariante no tempo, e expressa em termos da precisão ϕ , isto é $h_t = h = \frac{1}{\phi}$. Todas as variâncias na definição do modelo se admitem multiplicadas pelo factor de escala h . Ter-se-á :

$$\text{Equação das observações} \quad : y_t = f_t \theta_t + \epsilon_t \quad ; \epsilon_t \sim N(0, h = \frac{1}{\phi}) \quad (9.19)$$

$$\text{Equação de transição} \quad : \theta_t = G_{t-1} \theta_{t-1} + R_t \eta_t \quad ; \eta_t \sim N_q(0, h Q_t^*) \quad (9.20)$$

$$\text{Distribuições à priori} \quad : (\theta_0 | D_0) \sim N_m(a_0, h P_0^*) \quad (9.21)$$

$$: (\phi | D_0) \sim G\left(\frac{\alpha_0}{2}, \frac{\beta_0}{2}\right) \quad (9.22)$$

Como se provou em (8.2.2) quando a média de um processo Normal independente é conhecida, mas a precisão é considerada uma variável aleatória, a distribuição conjugada natural à priori é uma Gama, isto é,

$$f(\phi | D_0) \equiv f(\phi, \alpha_0, \beta_0) \propto \phi^{\frac{1}{2}\alpha_0 - 1} \exp\left\{-\frac{1}{2}\beta_0 \phi\right\}, \quad (\phi > 0)$$

Sendo a constante de proporcionalidade $(\beta_0/2)^{\frac{\alpha_0}{2}} \frac{1}{\Gamma(\alpha_0/2)}$, onde $\Gamma(\cdot)$ é a função Gama, ou seja, como é sabido da teoria das distribuições de probabilidade, $\beta_0 \phi$ tem uma distribuição Qui-quadrado com α_0 graus de liberdade,

$$\beta_0 \phi \sim \chi_{\alpha_0}^2.$$

Note-se que as quantidades iniciais a_0 , P_0^* , α_0 e β_0 devem ser especificadas, bem como as matrizes Q_t^* . Para a especificação de α_0 e β_0 pode ajudar o facto de que $E[\phi | D_0] = \alpha_0/\beta_0 = 1/s_0$, onde s_0 é uma estimativa pontual à priori da variância das observações h .

Como se viu no modelo Normal com todos os parâmetros conhecidos, todas as distribuições à priori e à posteriori são Normais. Assim aqui as distribuições condicionadas a h são também Normais, tendo-se ,

$$(\theta_{t-1} | D_{t-1}, h) \sim N_m(a_{t-1}, h P_{t-1}^*) \quad (9.23)$$

$$(\theta_t | D_{t-1}, h) \sim N_m(a_{t|t-1}, h P_{t|t-1}^*) \quad (9.24)$$

$$(y_t | D_{t-1}, h) \sim N(\bar{y}_{t|t-1}, h z_{t|t-1}^*) \quad (9.25)$$

$$(\theta_t | D_t, h) \sim N_m(a_{t|t}, h P_{t|t}^*) \quad (9.26)$$

com,

$$a_{t|t-1} = G_{t-1} a_{t-1}$$

$$P_{t|t-1}^* = G_{t-1} P_{t-1}^* G_{t-1}^T + R_t Q_t^* R_t^T$$

$$\bar{y}_{t|t-1} = f_t a_{t|t-1}$$

$$z_t^* = 1 + f_t P_{t|t-1}^* f_t^T$$

$$P_{t|t}^* = P_{t|t-1}^* - A_t A_t^T z_t^*$$

$$a_{t|t} = a_{t|t-1} + A_t \nu_t \quad ; \quad \nu_t = y_t - \bar{y}_{t|t-1} \quad \text{e} \quad A_t = P_{t|t-1}^* f_t / z_t^*$$

Quanto à distribuição à posteriori de ϕ como se está a trabalhar com famílias de distribuições conjugadas à priori, já se sabe que terá de ser também uma Gama.

Com efeito assumindo que a distribuição à priori para a precisão ϕ no instante $t-1$ é uma Gama de parâmetros $(\alpha_{t-1}/2)$ e $(\beta_{t-1}/2)$,

$$f(\phi | D_{t-1}) \propto \phi^{\frac{1}{2}\alpha_{t-1} - 1} \exp \left\{ -\frac{1}{2} \beta_{t-1} \phi \right\}, \quad (\phi > 0)$$

e como ,

$$(y_t | D_{t-1}, h = \phi^{-1}) \sim N(\bar{y}_{t|t-1}, \phi^{-1} z_{t|t-1}^*) \quad \text{ou seja}$$

$$f(y_t | D_{t-1}, h = \phi^{-1}) \propto \phi^{\frac{1}{2}} \exp \left[-\frac{1}{2z_t^*} \phi (y_t - \tilde{y}_{t|t-1})^2 \right] = \phi^{\frac{1}{2}} \exp \left(-\frac{\phi}{2} \cdot \frac{\nu_t^2}{z_t^*} \right)$$

Pelo teorema de Bayes temos que ,

$$f(\phi | D_t) \propto f(\phi | D_{t-1}) f(y_t | D_{t-1}, h = \phi^{-1})$$

$$\begin{aligned} f(\phi | D_t) &\propto \phi^{\frac{1}{2}\alpha_{t-1} - 1} \exp \left(-\frac{1}{2} \beta_{t-1} \phi \right) \phi^{\frac{1}{2}} \exp \left(-\frac{\phi}{2} \cdot \frac{\nu_t^2}{z_t^*} \right) = \\ &= \phi^{\frac{1}{2}(\alpha_{t-1}+1)-1} \exp \left[-\frac{1}{2}\phi \left(\beta_{t-1} + \frac{\nu_t^2}{z_t^*} \right) \right] \end{aligned}$$

Assim $(\phi | D_t) \sim G \left(\frac{\alpha_t}{2}, \frac{\beta_t}{2} \right)$ onde os parâmetros à posteriori são ,

$$\alpha_t = \alpha_{t-1} + 1 \quad \text{e} \quad \beta_t = \beta_{t-1} + \frac{\nu_t^2}{z_t^*}$$

O filtro de Kalman fornecendo em cada instante t a inovação ν_t e a respectiva variância z_t , permite a estimação sequencial da precisão ϕ e consequentemente a estimação da variância das observações h .

É também um resultado bem conhecido da teoria das probabilidades que as distribuições não condicionadas a h (desconhecido) são distribuições t de Student , com um número de graus de liberdade adequado. Assim ter-se-á ,

$$(\theta_{t-1} | D_{t-1}) \sim T_{\alpha_{t-1}}(a_{t-1}, P_{t-1}) \quad (9.27)$$

$$(\theta_t | D_{t-1}) \sim T_{\alpha_{t-1}}(a_{t|t-1}, P_{t|t-1}) \quad (9.28)$$

$$(y_t | D_{t-1}) \sim T_{\alpha_{t-1}}(\tilde{y}_{t|t-1}, z_{t|t-1}) \quad (9.29)$$

$$(\theta_t | D_t) \sim T_{\alpha_t}(a_{t|t}, P_{t|t}) \quad (9.30)$$

onde,

$$P_{t-1} = s_{t-1} P_{t-1}^*, \quad P_{t|t-1} = s_{t-1} P_{t|t-1}^*, \quad z_t = s_{t-1} z_t^* \quad \text{e} \quad P_{t|t} = s_t P_{t|t}^*$$

com $s_{t-1} = \frac{\beta_{t-1}}{\alpha_{t-1}}$ e $s_t = \frac{\beta_t}{\alpha_t}$ estimativas pontuais da variância das observações, respectivamente nos instantes $t-1$ e t .

A título de exemplo provaremos apenas para a distribuição preditiva (9.29).

Como $(y_t | D_{t-1}, h = \phi^{-1}) \sim N(\tilde{y}_{t|t-1}, \phi^{-1} z_{t|t-1}^*)$ e admitindo que,

$(\phi | D_{t-1}) \sim G(\frac{\alpha_{t-1}}{2}, \frac{\beta_{t-1}}{2})$, tem-se que a distribuição conjunta de y_t e ϕ é uma Normal - Gama,

$$\begin{aligned} f(y_t, \phi | D_{t-1}) &\propto \left(\frac{1}{2\pi z_{t|t-1}^*} \phi\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2z_{t|t-1}^*} \phi (y_t - \tilde{y}_{t|t-1})^2\right] \phi^{\frac{1}{2}(\alpha_{t-1})} \exp\left(-\frac{\phi}{2} \beta_{t-1}\right) \\ &= \phi^{\frac{1}{2}(\alpha_{t-1}-1)} \exp\left\{-\frac{1}{2} \phi \left[(y_t - \tilde{y}_{t|t-1})^2 / z_{t|t-1}^* + \beta_{t-1}\right]\right\} \end{aligned}$$

Assim, a distribuição marginal de y_t virá,

$$\begin{aligned} f(y_t | D_{t-1}) &= \int_0^\infty f(y_t, \phi | D_{t-1}) d\phi \quad \text{ou seja,} \\ f(y_t | D_{t-1}) &\propto \int_0^\infty \phi^{\frac{1}{2}(\alpha_{t-1}-1)} \exp\left\{-\frac{1}{2} \phi \left[(y_t - \tilde{y}_{t|t-1})^2 / z_{t|t-1}^* + \beta_{t-1}\right]\right\} d\phi = \\ &= \frac{\Gamma(\frac{\alpha_{t-1}+1}{2})}{\frac{1}{2} \left[(y_t - \tilde{y}_{t|t-1})^2 / z_{t|t-1}^* + \beta_{t-1}\right]^{\frac{1}{2}(\alpha_{t-1}+1)}} \quad , \text{ onde } \Gamma(a) = \alpha^a \int_0^\infty e^{-\alpha x} x^{a-1} dx . \end{aligned}$$

Assim,

$$f(y_t | D_{t-1}) \propto \left[\alpha_{t-1} + \frac{(y_t - \tilde{y}_{t|t-1})^2}{\frac{\beta_{t-1}}{\alpha_{t-1}} z_{t|t-1}^*} \right]^{-\frac{\alpha_{t-1}+1}{2}} ,$$

temos assim o núcleo de uma distribuição t univariada com α_{t-1} graus de liberdade, moda $\tilde{y}_{t|t-1} = f_t \hat{a}_{t|t-1}$, e factor de escala $\frac{\beta_{t-1}}{\alpha_{t-1}} z_{t|t-1}^* = s_{t-1} z_{t|t-1}^*$, como se pretendia provar.

As equações de actualização do filtro de Kalman são :

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{A}_t \nu_t \quad (9.31)$$

$$\mathbf{P}_{t|t} = s_t \mathbf{P}_{t|t}^* = s_t [\mathbf{P}_{t|t-1}^* - \mathbf{A}_t \mathbf{A}_t^T \mathbf{z}_t^*] = \frac{s_t}{s_{t-1}} [\mathbf{P}_{t|t-1} - \mathbf{A}_t \mathbf{A}_t^T \mathbf{z}_t] \quad (9.32)$$

$$e \quad s_t = \frac{\beta_t}{\alpha_t}, \quad \alpha_t = \alpha_{t-1} + 1 \quad e \quad \beta_t = \beta_{t-1} + s_{t-1} \frac{\nu_t^2}{z_t} \quad (9.33)$$

Assim o " preço a pagar" pela inclusão de um estimador sequencial s_t é a transformação de todas as distribuições Normais condicionadas a h , em distribuições " t de Student" univariadas (para as observações) e multivariadas (para os parâmetros) quando o condicionamento ao parâmetro de precisão se suprime.

Uso de distribuições à priori não -informativas.

Considerando o MLD univariado com todas as componentes do vector de especificação conhecidas e as hipóteses (9.6) e (9.7), resta-nos o problema da definição da distribuição à priori para o vector de estado inicial θ_0 . Suponha-se que não existe informação à priori suficiente para a especificação de uma distribuição informativa e que se considera as distribuições iniciais de referência :

$$f(\theta_0 | D_0) \propto \text{constante} \quad (9.34)$$

$$f(\theta_1 | D_0) \propto \text{constante} \quad (9.35)$$

Vejamos, pois como evoluem as distribuições à priori e à posteriori de θ_t , isto é $f(\theta_t | D_{t-1})$ e $f(\theta_t | D_t)$, respectivamente.

A distribuição à posteriori é proporcional ao produto da verosimilhança pela distribuição à priori. A verosimilhança para a observação, y_t , é ,

$$f(y_t | \theta_t, D_{t-1}) \propto \exp \left[-\frac{1}{2} \phi \left(y_t - f_t \theta_t \right)^2 \right]$$

Assim, para $t=1$ tem-se,

$$f(\theta_1 | D_1) \propto \exp \left[-\frac{1}{2} \phi \left(y_1 - f_1 \theta_1 \right)^T \left(y_1 - f_1 \theta_1 \right) \right]$$

$$= \exp \left[-\frac{1}{2} (\theta_1^T \phi f_1^T f_1 \theta_1 - 2 \theta_1^T \phi f_1^T y_1 + y_1^T y_1) \right]$$

Definindo $C_1 = \phi f_1^T f_1$ e $c_1 = \phi f_1^T y_1$ obtém-se ,

$$f(\theta_1 | D_1) \propto \exp \left\{ -\frac{1}{2} (\theta_1^T C_1 \theta_1 - 2 \theta_1^T c_1) \right\}$$

Para $t=2$ teremos,

$$f(\theta_2 | D_1) = \int f(\theta_2, \theta_1 | D_1) d\theta_1 = \int f(\theta_2 | \theta_1, D_1) f(\theta_1 | D_1) d\theta_1$$

Como $f(\theta_2 | \theta_1, D_1) \sim N_m (G_1 \theta_1, R_2 Q_2 R_2^T)$, designando genericamente $R_i Q_i R_i^T$ por W_i , para facilidade de notação, podemos escrever

$$f(\theta_2 | D_1) \propto \int \exp \left\{ -\frac{1}{2} [(\theta_2 - G_1 \theta_1)^T W_2^{-1} (\theta_2 - G_1 \theta_1) + \theta_1^T C_1 \theta_1 - 2 \theta_1^T c_1] \right\} d\theta_1$$

Desenvolvendo a expressão entre parentesis rectos temos,

$$\begin{aligned} & [(\theta_2 - G_1 \theta_1)^T W_2^{-1} (\theta_2 - G_1 \theta_1) + \theta_1^T C_1 \theta_1 - 2 \theta_1^T c_1] = \\ & = \theta_2^T W_2^{-1} \theta_2 - \theta_2^T W_2^{-1} G_1 \theta_1 - \theta_1^T G_1^T W_2^{-1} \theta_2 + \\ & + \theta_1^T G_1^T W_2^{-1} G_1 \theta_1 + \theta_1^T C_1 \theta_1 - 2 \theta_1^T c_1 = \\ & = \theta_1^T [C_1 + G_1^T W_2^{-1} G_1] \theta_1 - 2 \theta_1^T [c_1 + G_1^T W_2^{-1} \theta_2] + \theta_2^T W_2^{-1} \theta_2 = \end{aligned}$$

Definindo U_2 e α_2 tais que : $U_2 = C_1 + G_1^T W_2^{-1} G_1$ e $U_2 \alpha_2 = c_1 + G_1^T W_2^{-1} \theta_2$, vem

$$\begin{aligned} & = \theta_1^T U_2 \theta_1 - 2 \theta_1^T U_2 \alpha_2 + \theta_2^T W_2^{-1} \theta_2 = \\ & = (\theta_1^T - \alpha_2^T)^T U_2 (\theta_1^T - \alpha_2^T) + \theta_2^T W_2^{-1} \theta_2 - \alpha_2^T U_2 \alpha_2 \end{aligned}$$

Como $\int \exp \left\{ -\frac{1}{2} (\theta_1^T - \alpha_2^T)^T U_2 (\theta_1^T - \alpha_2^T) \right\} d\theta_1$ é uma constante em relação a θ_2 obtém-se ,

$$f(\theta_2 | D_1) \propto \exp \left\{ -\frac{1}{2} (\theta_2^T W_2^{-1} \theta_2 - \alpha_2^T U_2 \alpha_2) \right\}$$

Substituindo α_2 por $U_2^{-1} (c_1 + G_1^T W_2^{-1} \theta_2)$ e desenvolvendo $\theta_2^T W_2^{-1} \theta_2 - \alpha_2^T U_2 \alpha_2$ temos,

$$\begin{aligned} \theta_2^T W_2^{-1} \theta_2 - \alpha_2^T U_2 \alpha_2 &= \\ &= \theta_2^T W_2^{-1} \theta_2 - [U_2^{-1} (c_1 + G_1^T W_2^{-1} \theta_2)]^T U_2 [U_2^{-1} (c_1 + G_1^T W_2^{-1} \theta_2)] = \\ &= \theta_2^T [W_2^{-1} - W_2^{-1} G_1 U_2^{-1} G_1^T W_2^{-1}] \theta_2 - [c_1^T U_2^{-1} c_1 + \\ &+ c_1^T U_2^{-1} G_1^T W_2^{-1} \theta_2 + \theta_2^T W_2^{-1} G_1 U_2^{-1} c_1] = \\ &= \theta_2^T [W_2^{-1} - W_2^{-1} G_1 U_2^{-1} G_1^T W_2^{-1}] \theta_2 - 2 \theta_2^T W_2^{-1} G_1 U_2^{-1} c_1 - c_1^T U_2^{-1} c_1, \end{aligned}$$

Definindo $B_2 = W_2^{-1} - W_2^{-1} G_1 U_2^{-1} G_1^T W_2^{-1}$ e $b_2 = W_2^{-1} G_1 U_2^{-1} c_1$, podemos escrever,

$$\theta_2^T W_2^{-1} \theta_2 - \alpha_2^T U_2 \alpha_2 = \theta_2^T B_2 \theta_2 - 2 \theta_2^T b_2 - c_1^T U_2^{-1} c_1$$

Como $c_1^T U_2^{-1} c_1$ não depende de θ_2 chega-se a,

$$f(\theta_2 | D_1) \propto \exp \left\{ -\frac{1}{2} (\theta_2^T B_2 \theta_2 - 2 \theta_2^T b_2) \right\}$$

Pelo que a distribuição à posteriori de θ_2 ,

$$f(\theta_2 | D_2) \propto \text{verossimilhança} \times \text{dist. à priori}$$

$$\begin{aligned} f(\theta_2 | D_2) &\propto \exp \left\{ -\frac{1}{2} \phi^T (y_2 - f_2 \theta_2) (y_2 - f_2 \theta_2) \right\} \cdot \exp \left\{ -\frac{1}{2} (\theta_2^T B_2 \theta_2 - 2 \theta_2^T b_2) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [\theta_2^T (B_2 + \phi f_2^T f_2) \theta_2 - 2 \theta_2^T (b_2 + \phi f_2^T y_2)] \right\} \end{aligned}$$

Definindo $C_2 = B_2 + \phi f_2^T f_2$ e $c_2 = b_2 + \phi f_2^T y_2$, temos,

$$f(\theta_2 | D_2) \propto \exp \left\{ -\frac{1}{2} (\theta_2^T C_2 \theta_2 - 2 \theta_2^T c_2) \right\}$$

Assim, demonstrando por indução admita-se que para $t-1$ ($t > 1$), se tem

$$f(\theta_{t-1} | D_{t-2}) \propto \exp \left\{ -\frac{1}{2} \left(\theta_{t-1}^T B_{t-1} \theta_{t-1} - 2 \theta_{t-1}^T b_{t-1} \right) \right\}$$

$$f(\theta_{t-1} | D_{t-1}) \propto \exp \left\{ -\frac{1}{2} \left(\theta_{t-1}^T C_{t-1} \theta_{t-1} - 2 \theta_{t-1}^T c_{t-1} \right) \right\}$$

com ,

$$B_{t-1} = \mathbf{W}_{t-1}^{-1} - \mathbf{W}_{t-1}^{-1} \mathbf{G}_{t-2} \mathbf{U}_{t-1}^{-1} \mathbf{G}_{t-2}^T \mathbf{W}_{t-1}^{-1}$$

$$\mathbf{U}_{t-1} = \mathbf{C}_{t-2} + \mathbf{G}_{t-2}^T \mathbf{W}_{t-1}^{-1} \mathbf{G}_{t-2}$$

$$b_{t-1} = \mathbf{W}_{t-1}^{-1} \mathbf{G}_{t-2} \mathbf{U}_{t-1}^{-1} c_{t-2}$$

$$C_{t-1} = B_{t-1} + \phi \mathbf{f}_{t-1}^T \mathbf{f}_{t-1} \quad \text{e} \quad c_{t-1} = b_{t-1} + \phi \mathbf{f}_{t-1}^T y_{t-1}$$

Prove-se para $t=t$,

$$f(\theta_t | D_{t-1}) = \int f(\theta_t, \theta_{t-1} | D_{t-1}) d\theta_{t-1} = \int f(\theta_t | \theta_{t-1}, D_{t-1}) f(\theta_{t-1} | D_{t-1}) d\theta_{t-1}$$

Como $f(\theta_t | \theta_{t-1}, D_{t-1}) \sim N_m(\mathbf{G}_{t-1} \theta_{t-1}, \mathbf{W}_t)$, vem

$$f(\theta_t | D_{t-1}) \propto \int \exp \left\{ -\frac{1}{2} [(\theta_t - \mathbf{G}_{t-1} \theta_{t-1})^T \mathbf{W}_t^{-1} (\theta_t - \mathbf{G}_{t-1} \theta_{t-1}) + \theta_{t-1}^T C_{t-1} \theta_{t-1} - 2 \theta_{t-1}^T c_{t-1}] \right\} d\theta_{t-1}$$

$$\text{Como } f(\theta_t | \theta_{t-1}, D_{t-1}) = U_t / (C_{t-1} \mathbf{W}_t^{-1}) \text{ vem}$$

Desenvolvendo o argumento da exponencial e definindo U_t e α_t , tais que

$$U_t = C_{t-1} + \mathbf{G}_{t-1}^T \mathbf{W}_t^{-1} \mathbf{G}_{t-1} \quad \text{e} \quad U_t \alpha_t = c_{t-1} + \mathbf{G}_{t-1}^T \mathbf{W}_t^{-1} \theta_t$$

obtem-se tal como se fez para $t=2$,

Para que a função integranda na expressão (9.33) seja dependente de θ_t , logo a expressão (9.33) verifica-se,

$$f(\theta_t | D_{t-1}) \propto \exp \left\{ -\frac{1}{2} \left(\theta_t^T B_t \theta_t - 2 \theta_t^T b_t \right) \right\}$$

com, Como o vetor de estado é de dimensão m , portanto, das distribuições anteriores, a distribuição

$[y_t | x_t] \sim [y_t | x_t]$ se distribuições é gerada por uma distribuição pelo elemento de estado x_t e a distribuição

para y_t é dada por $y_t = \mathbf{G}_t x_t + v_t$, onde v_t é um vetor de dimensão m , portanto, a distribuição

para y_t é dada por $y_t = \mathbf{G}_t x_t + v_t$, onde v_t é um vetor de dimensão m , portanto, a distribuição

para y_t é dada por $y_t = \mathbf{G}_t x_t + v_t$, onde v_t é um vetor de dimensão m , portanto, a distribuição

$$B_t = \mathbf{W}_t^{-1} - \mathbf{W}_t^{-1} \mathbf{G}_{t-2} \mathbf{U}_t^{-1} \mathbf{G}_{t-1}^T \mathbf{W}_t^{-1}$$

$$b_t = \mathbf{W}_t^{-1} \mathbf{G}_{t-1} \mathbf{U}_t^{-1} c_{t-1}$$

A distribuição à posteriori de θ_t ,

$f(\theta_t | D_t) \propto$ verosimilhança \times dist. à priori

$$\begin{aligned} f(\theta_t | D_t) &\propto \exp \left\{ -\frac{1}{2} \phi \left(y_t - f_t \theta_t \right)^T \left(y_t - f_t \theta_t \right) \right\} \cdot \exp \left\{ -\frac{1}{2} \left(\theta_t^T B_t \theta_t - 2 \theta_t^T b_t \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\theta_t^T \left(B_t + \phi f_t^T f_t \right) \theta_t - 2 \theta_t^T \left(b_t + \phi f_t^T y_t \right) \right] \right\} \\ &\propto \exp \left[-\frac{1}{2} \left(\theta_t^T C_t \theta_t - 2 \theta_t^T c_t \right) \right] \end{aligned} \quad (9.36)$$

com, $C_t = B_t + \phi f_t^T f_t$ e $c_t = b_t + \phi f_t^T y_t$

Para completar a demonstração por indução temos de provar que para $t=1$ se tem,

$$f(\theta_1 | D_0) \propto \exp \left\{ -\frac{1}{2} \left(\theta_1^T B_1 \theta_1 - 2 \theta_1^T b_1 \right) \right\} \quad (9.37)$$

Partindo de $f(\theta_0 | D_0) \propto$ constante, temos que

$$f(\theta_1 | D_0) = \int f(\theta_1, \theta_0 | D_0) d\theta_0 = \int f(\theta_1 | \theta_0, D_0) f(\theta_0 | D_0) d\theta_0$$

Como $f(\theta_1 | \theta_0, D_0) \sim N_m(G_0 \theta_0, W_1)$, vem

$$\begin{aligned} f(\theta_1 | D_0) &\propto \int \exp \left\{ -\frac{1}{2} \left[(\theta_1 - G_0 \theta_0)^T W_1^{-1} (\theta_1 - G_0 \theta_0) + \theta_0^T C_0 \theta_0 - 2 \theta_0^T c_0 \right] \right\} d\theta_0 \\ &\propto \text{constante} \end{aligned} \quad (9.38)$$

visto que a função integranda na expressão (9.38) não depende de θ_1 , logo a expressão (9.37) verifica-se, bastando para isso considerar $B_1 = 0$ e $b_1 = 0$.

Como o vector de estado é de dimensão m , partindo das distribuições iniciais não-informativas (9.34) e (9.35) as distribuições à posteriori serão impróprias pelo menos até ao instante $t=m$. Depois do processamento de um número suficiente de observações as distribuições impróprias tornar-se-ão próprias. Assim, sendo n o número mínimo de observações necessárias para que as distribuições se tornem próprias,

$n = \min \{ t : \text{as distribuições à posteriori no instante } t \text{ são próprias} \}$

para $t \geq n$ é fácil concluir que (9.36) se identifica com (9.14) com $C_t = P_{t|t}^{-1}$ e $c_t = C_t a_{t|t}$, sendo portanto equivalentes as situações em que se parte de uma distribuição à priori não - informativa ou uma distribuição informativa Normal.

Com efeito, definindo a quantidade $a_{t|t} = C_t^{-1} c_t$ a expressão (9.36) pode escrever-se :

$$\begin{aligned} f(\theta_t | D_t) &\propto \exp \left\{ -\frac{1}{2} (\theta_t^T C_t \theta_t - 2 \theta_t^T c_t) \right\} = \\ &= \exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t})^T C_t (\theta_t - a_{t|t}) - c_t^T c_t \right\} \propto \\ &\propto \exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t})^T C_t (\theta_t - a_{t|t}) \right\} \end{aligned} \quad (9.39)$$

pois que $c_t^T c_t$ é constante relativamente a θ_t .

Considerando o caso em que a variância das observações h_t , não é conhecida, mas é invariante no tempo $h_t = h = \frac{1}{\phi}$, vimos que a distribuição conjunta à priori, não - informativa, para θ e h , aconselhável (Jeffreys (1961) e Bernardo (1979)) é,

$$f(\theta_1, h | D_0) \propto h^{-1} = \phi \quad (9.40)$$

Partindo desta distribuição à priori e considerando a matriz $W_t = h W_t^*$, seguindo um processo idêntico ao anterior, obter-se-iam as distribuições conjuntas, à priori e à posteriori, seguintes

$$f(\theta_t, h | D_{t-1}) \propto h^{-\left(\frac{\alpha_{t-1}}{2} + 1\right)} \exp \left\{ -\frac{1}{2} h^{-1} (\theta_t^T B_t^* \theta_t - 2 \theta_t^T b_t^* + \lambda_t) \right\} \quad (9.41)$$

$$f(\theta_t, h | D_t) \propto h^{-\left(\frac{\alpha_t}{2} + 1\right)} \exp \left\{ -\frac{1}{2} h^{-1} (\theta_t^T C_t^* \theta_t - 2 \theta_t^T c_t^* + \delta_t) \right\} \quad (9.42)$$

com ,

$$B_t^* = W_t^{*-1} - W_t^{*-1} G_{t-1} U_t^{*-1} G_{t-1}^T W_t^{*-1} \quad (9.43)$$

$$U_t^* = C_{t-1}^* + G_{t-1}^T W_t^{*-1} G_{t-1} \quad (9.44)$$

$$\mathbf{b}_t^* = \mathbf{W}_t^{*-1} \mathbf{G}_{t-1} \mathbf{U}_t^{*-1} \mathbf{c}_{t-1}^* \quad (9.45)$$

$$\mathbf{C}_t^* = \mathbf{B}_t^* + \mathbf{f}_t^T \mathbf{f}_t, \quad \mathbf{c}_t^* = \mathbf{b}_t^* + \mathbf{f}_t^T \mathbf{y}_t \quad (9.46)$$

$$\lambda_t = \delta_{t-1} - \mathbf{c}_{t-1}^* \mathbf{U}_t^{*-1} \mathbf{c}_{t-1}^* \quad (9.47)$$

$$\delta_t = \lambda_t + \mathbf{y}_t^2 \quad (9.48)$$

$$\alpha_t = \alpha_{t-1} + 1 \quad (9.49)$$

e valores iniciais $\mathbf{B}_1^* = \mathbf{0}$, $\mathbf{b}_1^* = \mathbf{0}$, $\lambda_1 = 0$ e $\alpha_0 = 0$

As distribuições à priori e à posteriori de θ_t , obtêm-se, respectivamente de (9.41) e (9.42), por marginalização.

Assim,

$$f(\theta_t | D_{t-1}) = \int_0^{+\infty} f(\theta_t, h | D_{t-1}) dh \propto \int_0^{+\infty} h^{-(\frac{\alpha_{t-1}}{2} + 1)} \exp \left\{ -\frac{1}{2} h^{-1} (\theta_t^T \mathbf{B}_t^* \theta_t - 2 \theta_t^T \mathbf{b}_t^* + \lambda_t) \right\} dh =$$

$$\int_0^{+\infty} \phi^{\left(\frac{\alpha_{t-1}}{2} - 1\right)} \exp \left\{ -\frac{1}{2} \phi (\theta_t^T \mathbf{B}_t^* \theta_t - 2 \theta_t^T \mathbf{b}_t^* + \lambda_t) \right\} d\phi$$

peço que se obtém ,

$$f(\theta_t | D_{t-1}) \propto [\lambda_t + \theta_t^T \mathbf{B}_t^* \theta_t - 2 \theta_t^T \mathbf{b}_t^*]^{-\frac{1}{2}(\alpha_{t-1} + 1)}$$

peço que a distribuição à priori de θ_t é uma distribuição t de Student com α_{t-1} graus de liberdade.

Analogamente se obteria ,

$$f(\theta_i | D_i) \propto [\delta_i + \theta_i^T C_i^* \theta_i - 2 \theta_i^T c_i^*]^{-\frac{1}{2}} (\alpha_i + 1)$$

sendo ainda a distribuição à posterior uma distribuição t com α_i graus de liberdade e estando os parâmetros da distribuição à posteriori relacionados com os parâmetros da distribuição à priori pelas relações (9.43) a (9.49).

9.2.2 ESPECIFICAÇÃO DA MATRIZ, Q_k - USO DE FACTORES DE DESCONTO .

Para completar a especificação do MLD bayesiano, resta-nos resolver o problema da especificação da matriz, Q_k , associada às perturbações do vector de estado, θ_k .

Muitos dos métodos de previsão assumem implicitamente que o conteúdo de informação de cada observação decresce com o passar do tempo.

O método de regressão exponencialmente ponderada (EWR), desenvolvido por Brown(1962) baseia-se num único factor de desconto " β " . Este método foi usado com sucesso em várias aplicações ; deixando, porém a desejar em situações mais complexas, por exemplo, quando existem diferentes fontes de variação sistemática no modelo. Assim, se um processo é caracterizado por uma tendência linear e sazonal, é de esperar que a taxa de perda de informação associada à parte sazonal seja bem mais suave do que a taxa associada à tendência . Em termos de factores de desconto, designando por β_1 e β_2 os factores de desconto da sazonalidade e da tendência respectivamente, deve ter-se : $\beta_1 < \beta_2$. Portanto é perfeitamente coerente considerar-se um vector de descontos em vez de um único factor de desconto para todas as componentes.

Ameen e Harrison (1984) definem o " método de estimação ponderada com desconto " (Discount Weighted Estimation - DWE) que é uma generalização dos mínimos quadrados ponderados ou regressão exponencialmente ponderada de Brown , considerando em vez de um único factor de desconto um vector ou matriz de desconto " B " . Ameen e Harrison (1985) introduzem o conceito de desconto nos modelos lineares de previsão bayesianos substituindo a equação de evolução do sistema (9.2) por uma relação descontada entre a precisão à posteriori associada a θ_{k-1} e a precisão à priori associada a θ_k .

Considerando o MLD estruturado de tal forma que a matriz G_k seja diagonal por blocos ,

$$G_k = \text{diag} \{ G_{k1}, G_{k2}, \dots, G_{ki} \} ,$$

onde o bloco G_{it} se admite ter dimensão $n_i \times n_i$ ($i=1,2,\dots,k$ e $\sum n_i = m$). A matriz de desconto B_i será ,

$B_i = \{ \beta_{1i} I_1, \dots, \beta_{ki} I_k \}$, onde $0 < \beta_{it} \leq 1$ e I_i é a matriz identidade de ordem n_i , para todo $i=1,2,\dots,k$. Assim o modelo Normal Bayesiano com desconto (Normal discount bayesian model - (N.D.B.M)) é formulado da seguinte maneira :

$$(\theta_{t-1} | D_{t-1}) \sim N_m (a_{t-1|t-1} ; P_{t-1|t-1}) , \quad (9.50)$$

$$(\theta_t | D_{t-1}) \sim N_m (a_{t|t-1} ; P_{t|t-1}) , \quad (9.51)$$

$$a_{t|t-1} = G_{t-1} a_{t-1|t-1} , \quad (9.52)$$

$$P_{t|t-1} = B_i^{-1/2} G_{t-1} P_{t-1|t-1} G_{t-1}^T B_i^{-1/2} , \quad (9.53)$$

$$e \quad (y_t | \theta_t) \sim N (f_t \theta_t ; h_t) , \quad (9.54)$$

A caracterização do MLD em cada instante t (que na formulação sem desconto era $M_t = \{ f_t, G_t, R_t, h_t, Q_t \}$ passa a ser $M_t = \{ f_t, G_t, R_t, h_t, B_t \}$ na formulação com desconto. A influência estocástica na evolução dos parâmetros não é directamente explicada através do ruído η_t . Esta será substituída por uma "relação guia" que estabelece apenas a evolução determinística de θ_t , ficando a aleatoriedade do processo por conta da matriz de desconto. A matriz de desconto B_i contém no máximo m elementos distintos, um para cada elemento do vector de estado. Note-se que alguns ou mesmo todos os factores de desconto podem coincidir, sendo, por isso, a especificação da matriz de desconto mais simples do que a matriz das variâncias-covariâncias, Q_t , para a qual em geral o utilizador tem menor sensibilidade. Para maior facilidade de exposição e de notação ,sem perda de generalidade, considere-se que a matriz R_i é a matriz identidade de ordem m .

West e Harrison (1989) introduzem o método de desconto por componentes (component discounting). Considerando o MLD geral composto pela sobreposição de vários sub-modelos ou componentes ,

$$M_i = \{ f_{it}, G_{it}, h_{it}, Q_{it} \} ,$$

com vectores de estado θ_{it} e perturbações aleatórias η_{it} para $i=1,2,\dots,k$ tal que :

$$\theta_{it} = G_{it-1} \theta_{it-1} + \eta_{it}$$

$$y_{it} = f_{it} \theta_{it} + \epsilon_{it}$$

$$\text{O MLD } y_i = f_i \theta_i + \epsilon_i \text{ e } \theta_i = G_{i-1} \theta_{i-1} + \eta_i \text{ com } y_i = \sum_0^{\infty} y_{it}, \epsilon_i = \sum_0^{\infty} \epsilon_{it} \text{ e } \eta_i = \sum_0^{\infty} \eta_{it}$$

e vector de estado de dimensão $m = n_1 + n_2 + \dots + n_k$, dado por ,

$$\theta_i^T = [\theta_{1i}^T \theta_{2i}^T \dots \theta_{ki}^T] , \text{ é assim especificado pelo quádruplo ,}$$

$$\{ f_i , G_i , h_i , Q_i \} , \text{ onde}$$

$$f_i = \{ f_{1i} , f_{2i} , \dots , f_{ki} \} ,$$

$$G_i = \text{diag} \{ G_{1i} , G_{2i} , \dots , G_{ki} \} =$$

$$= \begin{bmatrix} G_{1i} & 0 & 0 & \dots & 0 \\ 0 & G_{2i} & 0 & \dots & 0 \\ 0 & 0 & G_{3i} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & G_{ki} \end{bmatrix}$$

e

$$Q_i = \text{diag} \{ Q_{1i} , Q_{2i} , \dots , Q_{ki} \} =$$

$$= \begin{bmatrix} Q_{1i} & 0 & 0 & \dots & 0 \\ 0 & Q_{2i} & 0 & \dots & 0 \\ 0 & 0 & Q_{3i} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Q_{ki} \end{bmatrix}$$

A matriz \mathbf{Q}_t traduz um acréscimo de incerteza acerca do vector de estado entre os instantes $t-1$ e t . Mais precisamente, no instante $t-1$ a distribuição à posteriori para o vector de estado tem variância, $\text{var}[\theta_{t-1}|D_{t-1}] = \mathbf{P}_{t-1|t-1}$, que conduz através da equação de evolução, a uma variância para a distribuição à priori de θ_t , dada por,

$$\text{var}[\theta_t | D_{t-1}] = \mathbf{P}_{t|t-1} = \mathbf{G}_{t-1} \mathbf{P}_{t-1|t-1} \mathbf{G}_{t-1}^T + \mathbf{Q}_t$$

Designando por \mathbf{M}_t o primeiro termo isto é,

$$\mathbf{M}_t = \mathbf{G}_{t-1} \mathbf{P}_{t-1|t-1} \mathbf{G}_{t-1}^T = \text{var}[\mathbf{G}_{t-1} \theta_{t-1} | D_t]$$

pode considerar-se a variância à priori nula no modelo $\mathbf{M}_t = \{ \mathbf{F}_t, \mathbf{G}_t, \mathbf{h}_t, 0 \}$, ou seja o modelo standard sem ruído na equação de evolução. A adição do ruído aleatório η_t a $\mathbf{G}_{t-1} \theta_{t-1}$, para reproduzir a verdadeira equação de estado, tem o efeito já descrito de aumentar a incerteza ou perda de informação, traduzida por \mathbf{M}_t , na situação ideal, para a situação actual de $\mathbf{P}_{t|t-1} = \mathbf{M}_t + \mathbf{Q}_t$. Considerando um factor de desconto escalar β , ($0 < \beta \leq 1$) podemos admitir que a perda de informação ou aumento de incerteza se pode traduzir pela relação,

$$\beta \mathbf{P}_{t|t-1} = \mathbf{M}_t \text{ implicando assim que se tenha}$$

$$\mathbf{P}_{t|t-1} = \mathbf{M}_t + \mathbf{Q}_t \text{ com } \mathbf{Q}_t = \mathbf{M}_t(1-\beta)/\beta$$

Como já se referiu o uso de um só factor de desconto tem a implicação de que a informação decai à mesma taxa para todos os elementos do vector de estado. O que em geral não será adequado. Assim atendendo à estrutura por blocos assumida para a matriz \mathbf{G}_t podemos considerar idêntica estrutura para \mathbf{M}_t pelo que designando por \mathbf{M}_{it} ,

$$\mathbf{M}_{it} = \text{var}[\mathbf{G}_{it} \theta_{it} | D_{t-1}], \quad (i=1, \dots, k), \text{ podemos escrever,}$$

$\mathbf{P}_{it|t-1} = \mathbf{M}_{it} + \mathbf{Q}_{it}$, pelo que para cada sub-modelo ou componente \mathbf{M}_i podemos associar um factor de desconto β_i e definir $\mathbf{Q}_{it} = \mathbf{M}_{it}(1-\beta_i)/\beta_i$.

West e Harrison (1989) recomendam este método de desconto para quase todas as aplicações. É um método parcimonioso, pois aplica o mesmo factor de desconto a cada sub-modelo.

10 - CONTROLO E INTERVENÇÃO.

10.1 CONTROLO - Via factor de Bayes.

Os modelos bayesianos de previsão envolvem em geral o processamento sequencial das observações . A revisão progressiva das inferências e distribuições preditivas à medida que novas observações vão surgindo é crucial na abordagem bayesiana . O sucesso de um sistema de previsão exige flexibilidade no modelo para adequadamente detectar descontinuidades , no sistema gerador dos dados , como por exemplo, observações erráticas ou discrepantes ("outliers") , ou mudanças de estrutura . O principal objectivo consiste em detectar estas descontinuidades com a maior rapidez possível, permitindo assim que o modelo possa ser corrigido ou adaptado. O processo que passamos a descrever é o proposto por West (1986a) e West e Harrison (1986a) , baseado no factor de Bayes. Essencialmente o esquema de controlo consiste na comparação da capacidade preditiva do modelo em uso , com a de um modelo alternativo. Sendo este último construído sequencialmente, observação a observação e com a mesma estrutura do modelo que se está a usar. O problema principal na análise do desempenho de um modelo é o da construção do modelo alternativo adequado. De momento, admita-se que no instante t um modelo alternativo adequado existe. Designe-se o modelo em uso por M e o modelo alternativo por A . Em cada instante cada um dos modelos fornece uma distribuição preditiva para y_t dada a informação disponível D_{t-1} , respectivamente, $f_M(y_t | D_{t-1})$ e $f_A(y_t | D_{t-1})$. Como vimos na introdução desta segunda parte o factor de Bayes pode ser encarado como a razão de verosimilhanças, neste caso, da observação y_t dados respectivamente o modelo M e A , ou seja ;

$$B_t = \frac{f_M(y_t | D_{t-1})}{f_A(y_t | D_{t-1})} \quad (10.1)$$

Pequenos valores (< 1) de B_t indicam uma supremacia do modelo A sobre o modelo M , devendo este ser reexaminado .

Para um conjunto k de observações sucessivas , $y_t, y_{t-1}, \dots, y_{t-k+1}$ a " verosimilhança do modelo " corresponde à densidade preditiva conjunta ,

$$f(y_t, y_{t-1}, \dots, y_{t-k+1} | D_{t-1}) = f(y_t | D_{t-1}) f(y_{t-1}, \dots, y_{t-k+1} | D_{t-1}) .$$

Assim o factor de Bayes Global baseado em k observações consecutivas pode definir-se por :

$$\begin{aligned}
 B_t(k) &= \frac{f_M(y_t, y_{t-1}, \dots, y_{t-k+1} | D_{t-1})}{f_A(y_t, y_{t-1}, \dots, y_{t-k+1} | D_{t-1})} = \\
 &= \frac{f_M(y_t | D_{t-1}) f_M(y_{t-1}, \dots, y_{t-k+1} | D_{t-1})}{f_A(y_t | D_{t-1}) f_A(y_{t-1}, \dots, y_{t-k+1} | D_{t-1})} = B_t B_{t-1}(k-1)
 \end{aligned} \quad (10.2)$$

O factor de Bayes global ou cumulativo, $B_t(k)$, ou “ pesos de evidência” (Jeffreys(1961) , Good(1985) e West (1986a)) fornece a principal medida do desempenho ou capacidade preditiva do modelo M relativamente ao modelo alternativo A . Para cada k , $B_t(k)$ mede a evidência de M relativa a A , produzida pelas k observações mais recentes. Note-se que, a evidência a favor ou contra M se acumula multiplicativamente à medida que os dados são processados , isto é,

$$B_t(k) = B_t B_{t-1} \dots B_{t-k+1} \quad (k=1,2,\dots,t) \quad (10.3)$$

Note-se ainda que para $k=1$ se tem $B_t(1) = B_t$. Um valor pequeno para B_t é um aviso de que a observação y_t é possivelmente errática (“outlier”) ou o início de uma alteração de estrutura ; um pequeno valor de $B_t(k)$ para $k > 1$ sugere uma possível alteração no passado, pelo menos k períodos atrás . Para determinar o ponto mais provável do possível início de alteração, é necessário identificar “ o mais discrepante ” grupo de recentes observações consecutivas. Ou seja determinar o mínimo de $B_t(k)$ relativamente a k. Seja ,

$$V_t = \min_{1 \leq k \leq t} B_t(k) \quad (10.4)$$

com $V_1 = B_1$. As quantidades V_t podem ser calculadas sequencialmente pela relação

$$V_t = B_t \min \{ 1, V_{t-1} \} \text{ para } t=2, 3, \dots . \text{ Com efeito } B_t(1) = B_t \text{ e}$$

$$B_t(k) = B_t B_{t-1}(k-1) \text{ para } 2 \leq k \leq t, \text{ então ,}$$

$$\begin{aligned}
 V_t &= \min \{ B_t, \min_{2 \leq k \leq t} B_t B_{t-1}(k-1) \} = B_t \min \{ 1, \min_{2 \leq k \leq t} B_{t-1}(k-1) \} = \\
 &= B_t \min \{ 1, \min_{1 \leq j \leq t-1} B_{t-1}(j) \} = B_t \min \{ 1, V_{t-1} \}
 \end{aligned}$$

O mínimo no instante t é obtido para $K = l_t$, com $V_t = B_t(l_t)$, sendo os inteiros l_t sequencialmente atualizados por ,

$$l_t = \begin{cases} 1+l_{t-1} & \text{se } V_{t-1} < 1 \\ 1 & \text{se } V_{t-1} \geq 1 \end{cases}, \text{ pois que } V_t = B_t \text{ se e só se } l_t = 1, \text{ caso contrário } V_t = B_t V_{t-1}$$

e $l_t = 1 + l_{t-1}$. l_t é designado por "comprimento da descontinuidade" e fornece uma indicação do ponto mais provável do início da alteração. O conjunto de observações $y_t, y_{t-1}, \dots, y_{t-l_t+1}$ é o grupo de observações mais discrepante de interesse.

A sequência $\{V_t\}$ constitui assim, um monitor sequencial do desempenho de M relativamente a A. Na prática o factor de Bayes cumulativo, V_t , é usado como um teste estatístico sequencial de razão de probabilidades (SQRT - Sequential probability ratio test) (veja-se por exemplo Berger (1985) cap.7). O modelo M é considerado sob controlo até que haja evidência da sua "falha", medida por um valor pequeno de V_t . É pois necessário fixar um valor limite τ , tal que se $V_t < \tau$, o modelo é considerado "fora de controlo" e o analista deverá intervir, corrigindo ou adaptando o modelo. Os valores de τ variam entre 0 e 1 aconselhando West e Harrison (1989), valores entre 0.1 e 0.2. Se um único ponto discrepante é um valor errático ("outlier") deverá ser ignorado e o modelo adaptado tão depressa quanto possível de modo a que previsões futuras não sejam afectadas. Veremos adiante como tratar deste problema. Se o ponto discrepante é o início de uma alteração de estrutura a sua rejeição pode implicar a perda de informação importante, não devendo portanto ser feita.

O seguinte esquema lógico é um guia para o uso do factor de Bayes na detecção e rejeição de "outliers". No instante t , independentemente do que tenha acontecido antes, West (1986a) sugere que se adopte a seguinte regra de decisão:

(1) Se $V_t \geq \tau$ considere-se o sistema sobre controlo.

(2) Se $V_t < \tau$ então:

(i) se $l_t = 1$ a observação y_t é um potencial "outlier" e deverá ser omitida da análise, sendo tratada como uma observação em falta.

(ii) se $l_t > 1$ possível mudança de estrutura que ocorreu provavelmente a partir de $t - l_t + 1$.

(iii) se $l_t > k$, onde k é um inteiro pré-estabelecido, mesmo que $V_t \geq \tau$, podemos também admitir que é provável que uma mudança estrutural lenta esteja ocorrendo, provavelmente a partir de $t - l_t + 1$.

10.1.1 Construção de um Modelo Alternativo .

Na construção de um modelo alternativo A, segundo West (1986) é aconselhado que a densidade preditiva $f_A (y_t | D_{t-1})$ tenha uma localização e forma semelhantes às de $f_M (y_t | D_{t-1})$, embora sendo mais difusa. Então a construção da distribuição preditiva para o modelo A deve ser feita pelo mesmo processo que para o modelo M, isto é marginalizando a densidade conjunta $(y_t, \theta_t | D_{t-1})$,

$$f_A (y_t | D_{t-1}) = \int_{\Theta} f(y_t | \theta_t) f_A (\theta_t | D_{t-1}) d\theta_t ,$$

onde a densidade à priori alternativa, $f_A (\theta_t | D_{t-1})$, tem a mesma forma funcional que a correspondente densidade à priori para o modelo M, semelhante localização, mas maior dispersão.

West (1986) constrói $f_A (\theta_t | D_{t-1})$ baseando-se no uso do factor de desconto em potência (power discounting). Assim para $0 < \delta_t < 1$, a densidade à priori alternativa será:

$$f_A (\theta_t | D_{t-1}) \propto \left[f_M (\theta_t | D_{t-1}) \right]^{\delta_t} \quad (10.5)$$

O factor δ_t desconta efectivamente a informação contida na distribuição à priori para o modelo M produzindo uma alternativa com uma forma mais achatada, isto é com maior dispersão. Com efeito, por exemplo, se $(\theta_t | D_{t-1}) \sim N(a_{t|t-1}, P_{t|t-1})$ então,

$$f_M (\theta_t | D_{t-1}) \propto \exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t-1})^T P_{t|t-1}^{-1} (\theta_t - a_{t|t-1}) \right\}$$

pelo que

$$\begin{aligned} f_A (\theta_t | D_{t-1}) &\propto \left[\exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t-1})^T P_{t|t-1}^{-1} (\theta_t - a_{t|t-1}) \right\} \right]^{\delta_t} = \\ &= \exp \left\{ -\frac{1}{2} (\theta_t - a_{t|t-1})^T P_{t|t-1}^{-1} \delta_t (\theta_t - a_{t|t-1}) \right\} \end{aligned}$$

tem a mesma localização, mas matriz de variâncias-covariâncias $P_{t|t-1} / \delta_t$.

West (1986) estabelece algumas normas práticas para a determinação apropriada de δ_t , em geral valores no intervalo (0.1, 0.4) serão apropriados.

10.2 INTERVENÇÃO

10.2.1 Tratamento de uma observação, y_t , errática.

Diversas circunstâncias podem ocorrer que conduzam a que uma simples observação, y_t , assumam um valor discrepante sem qualquer relação com o resto da série. Embora a informação contida nessa observação possa ser importante em si, porém não deverá ser usada na actualização do modelo para o efeito de previsão, pois que para este efeito não conterá informação relevante. Assim a observação deverá ser omitida da análise, tratando-a como se na realidade estivesse em falta. Considerando a informação

$$I_t = \{ y_t, \text{ não está disponível} \}, \text{ teremos } D_t = \{ I_t, D_{t-1} \} \equiv D_{t-1}$$

pelo que a distribuição à posteriori para o vector de estado no instante t é igual à distribuição à priori, não sofrendo os parâmetros quaisquer actualização. Para maior facilidade de exposição considerar-se-á o caso em que se admite a variância das observações conhecida, pelo que todas as distribuições à priori, à posteriori e distribuição preditiva são Normais. Aplicando-se, naturalmente, os resultados ao caso da variância desconhecida, substituindo como vimos, as distribuições Normais por distribuições t de Student. Assim teremos,

$$(\theta_t | D_t) \equiv (\theta_t | D_{t-1}) \sim N(a_{t|t-1}; P_{t|t-1}).$$

Formalmente, esta situação pode ser introduzida no MLD de diversas maneiras, a mais simples e talvez mais apropriada, é precisamente encarar a observação, y_t , como tendo uma variância, h_t (H_t no caso multivariado) muito grande, considerando-se que h_t tende para infinito ou equivalentemente h_t^{-1} tende para zero. Nas equações de actualização a variância do erro de previsão, z_t , (Z_t no caso multivariado) tende para infinito com h_t . Assim formalmente ter-se-á,

$$I_t = \{ h_t^{-1} = 0 \}.$$

É natural que a omissão da observação y_t traga maior incerteza acerca das componentes de θ_t , pelo que será aconselhável intervir ao nível da evolução de θ_t , de modo a ter em conta o possível acréscimo da variância à posteriori de θ_t , $P_{t|t-1}$.

10.2.2 Perturbação aleatória adicional na equação de evolução.

O acréscimo de incerteza sobre as componentes de θ_t pode ser introduzido no modelo de uma maneira muito simples, bastando acrescentar à equação de evolução de θ_t uma nova componente aleatória, ζ_t , de média u_t e matriz de variâncias-covariâncias U_t ,

$$\zeta_t \sim N(u_t, U_t).$$

Admitindo-se que ζ_t é não correlacionado com $(\theta_{t-1} | D_{t-1})$ e com η_t , pelo que é também não correlacionado com $(\theta_t | D_{t-1})$. Assim, traduzindo a intervenção pela informação $I_t = \{u_t, U_t\}$ e substituindo no modelo a equação de transição por,

$$\theta_t = G_{t-1} \theta_{t-1} + \eta_t + \zeta_t \quad (10.6)$$

ter-se-á,

$$(\theta_t | I_t, D_{t-1}) \sim N[a_{t|t-1}^*, P_{t|t-1}^*], \quad (10.7)$$

onde

$$a_{t|t-1}^* = a_{t|t-1} + u_t \quad \text{e} \quad P_{t|t-1}^* = P_{t|t-1} + U_t. \quad (10.8)$$

Alguns dos elementos de u_t , bem como de U_t , podem ser nulos, permitindo assim a não alteração de algumas das componentes de θ_t , para as quais não se anteveja alteração.

A especificação de u_t e U_t em geral não é fácil. Quanto a u_t , em muitas situações a média $a_{t|t-1}^*$ é especificada sendo u_t determinado por $u_t = a_{t|t-1}^* - a_{t|t-1}$. Para U_t em geral é usado o mesmo processo de desconto que foi descrito para a especificação de Q_t . Assim se Q_t foi estruturada usando um conjunto standard de factores de desconto, aplicando esse mesmo conjunto a U_t , o MLD condicionado à informação $\{I_t, D_{t-1}\}$ pode ser reformulado escrevendo agora a equação de evolução como:

$$\theta_t = G_{t-1} \theta_{t-1} + \eta_t^*, \quad (10.9)$$

onde

$$\eta_t^* = \eta_t + \zeta_t \quad \text{e} \quad \eta_t^* \sim N[u_t, Q_t^*] \quad \text{com} \quad Q_t^* = Q_t + U_t. \quad (10.10)$$

Outros modos de intervenção poderiam ser considerados, bem como o aspecto da conjugação de mais de uma intervenção no instante t , porém não o faremos para não tornar demasiado extensa esta exposição.

West e Harrison (1989, cap11) apresentam várias aplicações onde ilustram diferentes tipos de intervenção nos modelos bayesianos.

11. CONCLUSÃO.

A metodologia da representação de modelos sob a forma de **espaço de estados** constitui um ponto de partida comum às abordagens clássica e bayesiana dos métodos de previsão.

O modelo linear dinâmico (MLD) permite modelizar fenómenos económicos e sociais onde a variabilidade dos parâmetros e a aleatoriedade da evolução são permanentes, não sendo, portanto requerida a **estacionaridade**.

O filtro de Kalman é um processo de estimação eficiente, que enquadra estimadores clássicos e bayesianos como seus casos particulares. O filtro de Kalman pode dizer-se que desempenha um papel notável como método unificador dos diferentes processos estatísticos de previsão e é aplicável a qualquer modelo desde que seja possível a sua representação sob a forma de espaço de estados.

A ampla classe de **modelos estruturais** constitui uma alternativa à classe dos modelos ARIMA introduzidos por Box e Jenkins. A principal diferença reside na metodologia de selecção do modelo.

Um sistema de previsão não se deve limitar apenas ao método estatístico mas, permitir também a interacção do utilizador. Os modelos estruturais de previsão bayesianos parecem poder incorporar de uma forma simples os conhecimentos que os analistas e mesmo outras pessoas ligadas ao uso das previsões detêm relativamente ao comportamento do fenómeno, assim como ao meio envolvente subjacente ao estudo a efectuar.

A abordagem bayesiana de previsão apresenta várias atracções, particularmente o aspecto de poder ser aplicada para pequenas amostras, ou mesmo na ausência de observações. De notar que a não existência de observações, não quer dizer ausência de informação.

Os métodos de previsão bayesiana constituem um processo de aprendizagem sequencial, que aliado à possibilidade de intervenção, assim que informação relevante sobre o processo surja, lhe conferem grandes potencialidades.

O problema de observações em falta pode ser facilmente tratado pelo filtro de Kalman bayesiano, simplesmente omitindo as correspondentes equações de actualização, isto é, considerando as distribuições à posteriori iguais às distribuições à priori. E uma vez tendo sido processadas todas as

observações o processo de alisamento pode fornecer estimativas das observações em falta.

O problema da existência de observações erráticas ("outliers"), bem como o de alterações estruturais podem ser automaticamente detectados e controlados - via factor de Bayes.

Em termos de comparação entre as abordagens clássica e bayesiana podemos dizer que a formulação matemática é a mesma, residindo a principal diferença ao nível da especificação dos hiperparâmetros, ou seja, relativamente ao seu processo de estimação. Na abordagem clássica é usado o método de estimação de máxima verosimilhança, sendo esta escrita em termos das inovações, o que em geral requer um número de observações considerável. Na abordagem bayesiana os hiperparâmetros dos modelos são designados subjectivamente, ou como no caso das variâncias via " factores de desconto".

Parece-nos no entanto que os métodos bayesianos, permitindo uma maior facilidade de intervenção, serão em geral mais adequados. No entanto, como é óbvio, não existe para todas as situações um método "melhor" que qualquer outro.

A escolha do método dependerá de vários factores, tais como, o objectivo do estudo, o tipo de dados, o número de observações disponíveis, a existência ou não de informação à priori, o horizonte de previsão, a disponibilidade de programas informáticos capazes de tratar os modelos, etc.

Lamenta-se bastante que, devido a alguns problemas de "software", não tenha sido possível completar este trabalho com algumas aplicações. Pois que uma das razões que motivaram a escolha do tema foi precisamente a possibilidade de vir a efectuar uma parte prática.

No que diz respeito à abordagem clássica, dispunha-se do programa STAMP (Structural Time Series Analyser Modeller and Predictor), porém para a análise bayesiana, o programa BATS (Bayesian Analysis Time Series) desenvolvido por West, Harrison e Pole (1987), revelou-se com alguns problemas que não nos foi possível ultrapassar.

Dos vários pontos importantes que ficaram por analisar mencionaremos apenas alguns, tais como :

- problemas de selecção ou identificação de modelos,
- testes e medidas de ajustamento,
- modelos não lineares e não gaussianos,
- combinação de modelos (multi-process models).

BIBLIOGRAFIA

- Aitchison, J. e Dunsmore, I. R. (1975). Statistical Prediction Analysis. Cambridge University Press, Cambridge.
- Ameen, J. R. M., e Harrison, P.J. (1984). Discount Weighted Estimation. J. of Forecasting 3, 285 - 296.
- Ameen, J. R. M., e Harrison, P.J. (1985a). Normal Discount Bayesian Models. In Bayesian Statistics 2, 271-298, J.M. Bernardo, M.H. DeGroot, D.V. Lindley e A.F.M. Smith (Eds.). North- Holland, Amsterdam, e Valencia University Press.
- Ameen, J. R. M., e Harrison, P.J. (1985b). Discount Bayesian Multi - Process Modelling with Cusums. In Time Series Analysis : Theory and Practice 5, O.D. Anderson (Ed.). North- Holland, Amsterdam.
- Anderson, B.D.O., e Moore, J.B. (1979). Optimal Filtering. Prentice - Hall, New - Jersey.
- Ansley, C.F. (1979). An Algorithm for Exact Likelihood of a Mixed Autoregressive Moving Average Process. Biometrika 66, 59-65.
- Ansley, C.F. e Kohn, R. (1985). Estimation, Filtering, and Smoothing in State Space Models with Incompletely specified Initial Conditions. The Annals of Statistics, 13, n° 4, 1286-1316
- Aoki, M. (1987). State Space Modelling of Time Series. Springer - Verlag, Berlin Heidelberg.
- Barnett, V. (1982). Comparative Statistical Inference. John Wiley & Sons.
- Berger, J.O., (1985). Statistical Decision Theory and Analysis (2nd edn.). Springer - Verlag, New York.
- Bernardo, J. M. (1979). Reference Posterior Distribution for Bayesian Inference (with discussion). J. Roy. Statist. Soc. (Ser. B) 41, 113 - 148.
- Box, G. E. P. (1980). Sampling and Bayes' Inference in Scientific Modelling and Robustness. J. Roy. Statist. Soc. (Ser. A) 143, 383-430.
- Box, G. E. P. (1983). An Apology for Ecumenism in Statistics. In Scientific Inference, Data Analysis and Robustness, Ed. G. E. P. Box, Tom Leonard e Chien- FuWu, Academic Press, New York.
- Box, G. E. P. (1984). The Importance of Practice in the Development of Statistics. Technometrics, 26, 1-8.
- Box, G. E. P., e Tiao G. C. (1968). A Bayesian Approach To Some Outlier Problems. Biometrika 55, 119-129
- Box, G. E. P., Pierce, D.A. (1970). Distribution of Residual Autocorrelation in Autoregressive

- Integrated Moving Average Time Series Models. *J. Amer. Statist. Ass.* **65**, 1509 - 1526.
- Box, G. E. P., e Tiao G. C. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *J. Americ. Statist. Ass.* , **70** , 70- 79.
- Box, G. E. P. , Jenkins, G.M. (1970/76). *Time Series Analysis : Forecasting and Control*. Holden- Day, San Francisco.
- Box, G. E. P., Hilmer S. C. e Tiao G. C. (1978). Analysis and Modelling of Seasonal Time Series . In A. Zellner (Ed.) *Seasonal Analysis of Economic Time Series*, 309-334. Washington , D.C. : U S Department of Commerce.
- Brockwell, P. J. Davis, R.A. (1987) . *Time Series: Theory and Methods*. Springer - Verlag, New York.
- Broemling, L.D.(1985). *Bayesian Analysis of Linear Models*. Marcel Dekker, New York.
- Brown , R.G. (1962). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice - Hall, Englewood Cliffs, NJ.
- Crowder, M. J. (1976) . Maximum Likelihood Estimation of Dependent Observations. *J. Roy. Statist. Soc. (ser. B)* **38** , 45-53
- de Jong, P. (1988). The Likelihood for a State Space Model. *Biometrika* **75** , 165 -169.
- Diaconis, P.e Ylvisaker, D.(1985). Quantifying Prior Opinion. In *Bayesian Statistics 2*, 133-156 , J.M. Bernardo, M.H. DeGroot, D.V. Lindley e A.F.M.Smith (Eds.).North- Holland ,Amsterdam, e Valencia University Press.
- Engle, R.F.(1978) . Estimating Structural Models of Seasonality. In A. Zellner (Ed.). *Seasonal Analysis of Economic Time Series*. Washington D. C. Bureau of the Census.
- Gersch, W. , Kitagawa, G. (1983) . The prediction of Time Series with Trends and Seasonalities. *J. Business & Economic*, **1**, n°3 , 253 -264.
- Gilchrist, W. G. (1967). Methods of Estimation Involving Discounting. *J. Royal Statistical Soc.* **29** n° 2, 235-369
- Goldstein, M (1976). Bayesian Analysis of Regression Problems. *Biometrika* **63**, 51 - 58.
- Good, I.J., (1985). Weights of Evidence : a critical Survey. In *Bayesian Statistics 2*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley e A.F.M. Smith (Eds.). North - Holland, Amsterdam, e Valencia University Press.
- Granger, C. W. J., e Ramanathan, R. (1984). Improved Methods of Combinig Forecasts. *J. Forecasting* **3**, 197- 204
- Granger, C. W. J. e Newbold, Paul (1986). *Forecasting Economic Time Series*, (2nd Ed), Academic Press, London .
- Hannan, Turell e Tuckwell (1970) . The Seasonal Adjustment of Economic Time Series. *Int. Economic Review*, **11**, 24 - 52.

- Harrison, P. J. (1965). Short - Term Sales Forecasting. *Applied Statistics*, **15**, 102- 139.
- Harrison, P.J.(1988). Bayesian forecasting in O.R.. In *Developments in Operational Operational Research 1988*, N.B. Cook e A.M. Jonhson (Eds). Pergamon Press, Oxford.
- Harrison, P. J., e Stevens, C.F. (1971). A Bayesian Approach to Short-Term Forecasting . *Oper. Res. Quart.* **22**, 341- 362.
- Harrison, P. J., e Stevens, C. F. (1976a) . Bayesian Forecasting (with discussion). *J.Royal Statistical Soc.. (ser. B)* **38**, 205 -247.
- Harrison, P. J., e Stevens, C. F. (1976b) . Bayes Forecasting in action : Case Studies . Warwick Research Report **14**, Department of Statistics, University of Warwick.
- Harrison, P.J. e Akram, M. (1983). Generalised Exponentially Weighted Regression and Parsimonious Dynamic Linear Modelling. In *Time Series Analysis : Theory and Practice 3*, O.D. Anderson (Ed.). North - Holland, Amsterdam.
- Harrison, P. J., e West, M. (1986). Bayesian Forecasting in practice. *Bayesian Statistics Study Year Report 13*, University of Warwick.
- Hartigan, J.A., (1969). Linear Bayesian methods. *J. Royal Statistical Soc. (Ser. B)* **31**, 446-454.
- Harvey, A.C. (1981) . The Econometric Analysis of Time Series. Philip Allan , U.K.
- Harvey, A.C. (1984a). A unified view of statistical forecasting procedures. *J.Forecasting* **3**, 245- 275.
- Harvey, A.C. (1984b/85) Trends and Cycles in Macroeconomic Time Series. *J. of Business & Economic Statistics* , **3**, n.º 3 , 216-227
- Harvey, A. C. (1989). Forecasting , Structural Time Series Models and the Kalman Filter. Cambridge University Press.
- Harvey , A.C. e Philips, G.D.A. (1979). The Estimation of Regression Models with Autoregressive - Moving Average Disturbances. *Biometrika*, **66**, **1**, 49- 58.
- Harvey, A.C. e P.H.J. Todd (1983). Forecasting Economic Time Series with Structural and Box - Box - Jenkins Models: A Case Study (with discussion). *Journal of Business and Economic Statistics*, **1** ,299-315.
- Harvey, A.C. e S.Peters, (1984). Estimation Procedures for Structural Time Series Models. LSE Econometrics Programme Discussion Paper A 44.
- Harvey, A.C., e J. Durbin (1986). The Effects of Seat Belt Legislation on British Road Road Casualties : A Case Study in Structural Time Series Modelling (with discussion). *J. Royal Statistical Soc. (Ser. A)* **149** ,187-227.

- Jaynes, E.T. (1985). Highly Informative Priors . In Bayesian Statistics 2, 329 - 360, J.M. Bernardo, M.H. DeGroot, D.V. Lindley e A.F.M. Smith (Eds.). North - Holland, Amsterdam, e Valencia University Press.
- Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory, Academic Press . New York.
- Jeffreys, H., (1961). Theory of Probability (3rd edn.). Oxford University Press, London.
- Johnson, N.L., e Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. John Wiley, New York.
- Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME, J. of Basic Engineering, Series 82 D, 35- 45.
- Kalman, R.E., Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. Transactions of the ASME, Journal of Basic Engineering, Series 83D, 95- 108
- Kitagawa, G., (1981). A Nonstationary Time Series Model and its Fitting By a Recursive Filter. J. of Time Series Analysis ,2 , n^o 2, 103-116.
- Kitagawa, G., (1987). Non- Gaussian State-Space Modelling of Nonstationary Time Series (with discussion) . J.Amer. Statist. Ass. 82, 1032-1063.
- Lindley, D.V. (1965) . Introduction to Probability and Statistics from a Bayesian Viewpoint .,vol 1 e 2, Cambridge University Press, Cambridge.
- Lindley, D.V., e Smith, A.F.M., (1972). Bayes' Estimates for the Linear Model. J. Roy. Statist. Soc. (Ser. B) 34, 1-41..
- Maravall, A. (1985). On Structural Time Series Models and The Characterization of Components. J. of Business & Economic Statistics, 3, n^o 4, , 350- 355.
- Meinhold, R. J., Singpurwala, N. D. (1983). Understanding the Kalman Filter. The American Statistician , 37 , n^o 2 , 123 - 127.
- Migon, H.S. e Harrison, P.J. (1985). An Application of Non-linear Bayesian Forecasting To Television Advertising. In Bayesian Statistics 2, 681 - 696, J.M. Bernardo, M. H. DeGroot D.V. Lindley e A.F.M. Smith (Eds.). North - Holland, Amsterdam, e Valencia University Press.
- Murteira, Bento J.F. (1988). Estatística : Inferência e Decisão. Imprensa Nacional- Casa da Moeda.
- Nerlove, P. (1974) . The Exact Likelihood Function for a Mixed Autorregressive - Moving Average Process. Biometrika , 61 , 423 - 426.
- Nerlove, P., D. M. Grether e J.L. Carvalho (1979). Analysis of Economic Time Series. Academic Press. New York.
- Priestley, M. B. (1981). Spectral Analysis and Time Series. Vol 1 e 2 , Academic Press, New York.
- Raiffa, H. e Schlaifer, R. (1961/72/77) . Applied Statistical Decision Theory. Harvard

University, Boston.

Renyi, A. (1961/70). Probability Theory. Amsterdam, North- Holland.

Roberts, S.A. e Harrison, P.J. (1984). Parsimonious Modelling and Forecasting of Seasonal Time Series. European J. Oper. Res. 16, 365- 377.

Rosenberg, B. (1973) . Random Coefficient Models : The Analysis of a Cross - Section of Time Series by Stochastically Convergent Parameter Regression . Annals of Economic and Social Measurement ,2 , 399 - 428.

S. James Press (1989). Bayesian Statistics Principles, Models, and Applications. John Wiley Applications. John Wiley & Sons . New York.

Sparkes, J.R. e Mchugh, A.K. (1984). Awareness and Use of Forecasting Techniques in British Industry. J. Of Forecasting, 3. 37-42.

West, M. (1981). Robust Sequential Approximate Bayesian Estimation. J. Roy. Statist. Soc. (Ser.B) 43, 157- 166.

West, M. (1986a) . Bayesian Model Monitoring. J. Roy.Statist.Soc. (Ser. B) 48, 70-8.

West, M., e Harrison, P.J., e Migon, H.S. (1985). Dynamic Generalised Linear Models and Bayesian Forecasting (with discussion). J. Amer. Statist. Ass. 80, 73- 97.

West, M., e Harrison, P.J. (1986). Monitoring and Adaptation in Bayesian Forecasting Models. J. Appl. Satist. Ass. ,81, 395 , 741- 750.

West, M., e Harrison , P.J. (1989) . Bayesian Forecasting and Dynamic Models. Springer- Verlag, New York.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley & Sons.